SOCIAL COGNITION: EFFECTS OF OUTCOME KNOWLEDGE AND ORDER OF INFORMATION ON JUDGEMENT UNDER UNCERTAINTY

DONALD C. PENNINGTON

Thesis submitted to the University of Warwick for the degree of Doctor of Philosophy, September, 1980.

for

I.E.M.

and

D.R.R.

without whom this thesis
would not have been possible.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

This thesis could not have been written without the assistance and encouragement of numerous people. To the many friends and colleagues who have borne with me during the research and writing periods I am greatly indebted. To single out a small number of people to thank will cause some omissions. To those not specifically mentioned please forgive this oversight.

Perhaps my greatest debt is to Ian Morley for not only being an excellent supervisor but for also being a friend. Both roles merged into one during the period of my research. My thanks for always being available and for calmly seeing me through the crises one inevitably meets when doing research.

To Derek Rutter a special word of appreciation. The encouragement and confidence in my ability saw me through moments of despair. Derek also inspired me with enthusiasm and led me to appreciate the importance of thorough experimental design (and not to split infinitives).

Professor Geoffrey Stephenson, who was to be my supervisor, I shall be eternally grateful to for providing the means to carry out the research.

My thanks to Professor John Annett for providing support and assistance whenever it was needed.

June, Lorraine and Peter I would particularly like to mention for putting up with me over the time I was engaged in this research.

George, Kyla and Tom, the latter appearing rather late on the scene, receive my deepest thanks. Bearing with me throughout the past three years was no easy task. This thesis is as much the product of George's endurance and encouragement as it is of my ability.

SUMMARY

This thesis investigates how (1) outcome knowledge; (2) information order influence judgement under uncertainty.

Previous research has indicated that when we know the outcome of some event (hindsight) perceptions are biased so that we see the outcome as more likely than when we do not have such knowledge (foresight). The major line of empirical enquiry in this thesis is to discover the conditions under which hindsight bias of this kind does and does not occur. In particular, an attempt is made to discover how information order effects hindsight bias.

The thesis is divided into three parts. In PART 1 the first chapter shows how hindsight bias is not an isolated phenomena but one which fits into a coherent body of research in cognitive social psychology. The second and third chapters critically review previous research on hindsight bias and information order respectively.

PART II details eight experiments carried out which investigate the effect of outcome knowledge and information order on different types of judgement under uncertainty. Experiments 1 - 3 investigated the effect of outcome knowledge and information order on subjective likelihoods. Experiment 4 examined information order with respect to juror judgements. Experiments 5 - 7 used material describing contemporary social/political events. Experiment 8 used factual material about sex roles.

These experiments are discussed in PART III of this thesis, which consists of two chapters. In the first chapter, Chapter 13, a general discussion is given to the findings of the eight experiments. The last chapter in this thesis considers cognitive processes that may be involved in hindsight judgements and concludes by considering the implications of this research for a model of man as an intuitive scientist.

Our hindsight drastically prunes the
possible paths events could have taken:
our foresight can never be so efficient.
(McCorduck, 1979, p.18).

CHAPTER 1.

INTRODUCTION

1.1. PERSPECTIVE

> "Perhaps the historian's most difficult handicap lies in his inevitable hindsight. It is true enough that only knowledge of what comes after makes him an historian, but that with which he can least dispense also most subtlely corrupts him. Since we know how things worked out we are twice tempted to suppose that they were bound to work out in this way, and consider the known outcome to be in someway 'right'. The first frees the historian of his basic duty to explain: if what happened was inevitable no thought is required in understanding it. The second makes him into a tedious defender of accomplished fact and leads him to consider the past only in the light of the present, the notorious weakness already denounced." (Elton, 1969, p99).

The above quote introduces an important theme of the following two chapters and of the empirical work reported in PART II of this thesis. People in possession of historical or outcome knowledge (hindsight), whether historians or lay people, have a privileged view of events -- they know what happened. Traditionally and intuitively such knowledge is believed to confer us with a "wisdom of hindsight". It is further believed that this knowledge allows us to make a more complete and insightful analysis of what went on than would otherwise be possible.

Recent research (Fischhoff, 1975a,b; Fischhoff & Beyth, 1975; Fischhoff, 1977; Slovic & Fischhoff, 1977; Wood, 1978) has seriously challenged this view, indicating that precisely the opposite may be the case. They found that judges given outcome knowledge, and thus operating in hindsight, perceived the outcome they believed to have occurred as more inevitable than judges who are operating in foresight and do not know how events turned out. This phenomena has become known as "hindsight bias" or the "knew-it-all-along" effect. It appears to confirm Elton's (1969) suspicions quoted at the beginning of this chapter. Historians, of course, are not the only group of people to view events from such a position. In everyday life people often attempt to understand events which have happened to them. Fischhoff's research, by the same token, would indicate that they too fail to benefit from the "wisdom of hindsight".

The importance and significance of this area of enquiry can be seen by taking some examples from international politics, and will serve to demonstrate why we need to systematically investigate the process of second guessing; that is the tendency for people to overestimate what they would have predicted in the same circumstances (Walster, 1967).

{1} Wohlstetter's(1962,1965) analysis of the Cuban Missile Crisis in the early sixties and the bombing of Pearl Harbour in 1941 shows that before each crisis occurred there was a wealth of information to

indicate that such outcomes would obtain. Why she asks:

> "with all this data didn't we know that Japan would attack Pearl Harbour on December 7th. Why, when it seems so clear in retrospect, didn't we anticipate that Khrushchev might put medium range missiles in Cuba? Why didn't we seize the first indications that such installations were on the way?" (Wohlstetter,1965,p694).

Recriminations have echoed down the years concerning why appropriate measures were not taken at the time to avoid the almost disastrous consequences in one case and the destruction of the Pacific Fleet in the other case. Were the decision-makers at the time to blame? If it was so obvious in hindsight, why was it not so in foresight?

{2} Correlli Barnett (1963), in his study of supreme command in the First World War, says that:

> "The statesmen of Europe....made frantic and genuine efforts in July (1914) to draw away from the slide into general war. It was all an illusion: they had no free will. Once Austria declared war on Serbia, Europe moved helplessly according to a deterministic fate" (p27).

Just how inevitable was this slide into general war? Was it perceived in this way by those in command before the war had become a fact, i.e. in foresight? If frantic and genuine efforts were made then one might think this not to be the case. Hermann (1972)discusses other research in social psychology relevant to this.

{3} Neville Chamberlain's appeasement policy to Hitler in the mid 1930's failed to curb Hitler's expansionism, and led to a German image of Britain as soft on the major issue of war. This led Hitler

to seriously miscalculate Britain's reaction to the invasion of Poland in 1939. In hindsight, historians have said that the appeasement policy could never have been successful and should have been realised at the time (Noguères, 1963). One effect of this was for future governments to regard appeasement policies with suspicion, with a tendency to regard them as inappropriate in any situation. Thus Anthony Eden's approach to the Suez Crisis of 1956 was one of aggression rather than appeasement. After the event historians have argued that appeasement would have been a much more appropriate and successful policy under the circumstances.

The first two examples highlight how after the fact analysis may lead to serious bias. For example, Fischhoff & Marom (1976) criticised Janis' (1972) analysis of Pearl Harbour on the grounds that he fails to acknowledge the limitations of retrospective analysis. Janis attributes the failure to predict the outcome as a product of faulty decision-making and information search resulting from a phenomena he calls "groupthink". Fischhoff & Marom point out that he may be subject to hindsight bias himself with the consequence that other important factors are ignored. That factors other than "groupthink" need to be taken into account is lucidly demonstrated by Wohlstetter (1962,1965). She argues that although America had an unprecedented intelligence picture of the enemy there was so much 'noise' in the system at the time that it was not feasible for such an eventuality to be seriously considered or predicted. Only in hindsight could the "true signals" be sorted out from the "noise".

The third example highlights a consequence of hindsight bias: our ability to learn from the past is impaired because, says Fischhoff (1975a):

> "When we attempt to understand past events, we implicitly test the hypotheses or rules we use both to interpret and anticipate the world around us. If, in hindsight, we systematically underestimate the surprises which the past held and holds for us, we are subjecting those hypotheses to inordinately weak tests and, presumably, finding little reason to change them. Thus, the outcome knowledge which gives us the feeling that we understand what the past was all about may prevent us from learning anything from it." (p298-299)

This is elaborated upon by May (1973) who shows that analogies from the past applied to current problems or crises may be inappropriate because they are oversimplified. Furthermore, in trying to formulate a policy of how to deal with the crises the analogy seized upon may be unduly influential. One consequence of this is that the outcome which occurred in the past, if undesirable, may be perceived as the inevitable result of an incorrect approach to the present problem. This results in the policy adopted in the past being seen as inappropriate to the current crisis only because it failed to bring about the desired result then. Such misuse of history abounds according to May (1973).

Ideally, one important task of anyone who makes use of historical knowledge, professional or lay person, is to provide an accurate reconstruction of events. This entails ignoring outcome knowledge and depicting things as they were in foresight. Logically there is no reason why this should not be achieved. In some sense it is demanded if we are to learn anything from the past and not be surprised by the

future. If, however, hindsight confers spurious wisdom upon us and encourages a knew-it-all-along attitude it is questionable just how useful such an analysis will be when we are trying to understand and predict events in the future.

In this thesis, I attempt to establish the degree to which outcome knowledge biases people's perceptions of social events. In addition, I attempt to discover some of the conditions which may encourage bias of this kind and some of the conditions which may serve to reduce or eliminate it. I have assumed that it will be important to consider the amount of information, type of information, and the order in which it is obtained. Consider, say, the task of an historian. Typically he first knows how some event turned out. In seeking to explain why this happened, or why some other outcome did not occur he will search for relevant information from what is available. The way this information is searched for and the order in which he reads it are likely to affect the explanation which he gives. Furthermore, outcome knowledge itself will partially determine what type of information is searched for, how it is assimilated, and the way in which it is judged relevant.

The order in which information describing some event appears may be an important variable and have a strong influence upon perceptions and judgments. This may be equally true of the professional historian and 'lay historian' interested in understanding or second guessing some past event. Consequently, the order in which relevant information is brought to a person's attention may have predictable and systematic influences upon how it is viewed.

In summary, then, this thesis has two themes running through it, one primary and one secondary. The primary theme is concerned with a detailed empirical investigation of how outcome knowledge (hindsight) affects people's perceptions and subjective likelihood assessments. In addition, some attempt is made to specify the conditions under which the bias will and will not occur. The secondary theme concerns order effects. This is empirically investigated both with respect to outcome knowledge and in juror decision-making.

1.2. ORGANISATION OF THIS THESIS.

Research into hindsight bias is one aspect of the more general study of the biases and inferential errors which people make. Much of this research treats man as an "intuitive scientist" and is central to the area known as cognitive social psychology.

Chapter 2 takes as its theme social psychology's view of rational man. The research reviewed in this chapter includes cognitive consistency; information search, assimilation and opinion revision; heuristics of thinking: bias in attribution, and, bias in prediction. This chapter serves two functions: it (1) attempts to show how a new paradigm to a rational model of man, man the intuitive scientist, is beginning to emerge. The intuitive scientist is characterised as using informal rules of inference and subject to serious biases; (2) demonstrates that the hindsight bias, briefly discussed in this chapter and treated at greater length in Chapters 2 and 3, is not an isolated phenomena. The chapter shows how this bias fits in and interrelates with other biases and inferential errors of the intuitive scientist.

Chapter 3 provides a detailed review of the empirical evidence investigating the effect of outcome knowledge and reporting hindsight bias. This chapter casts a critical eye over the six published studies and finds a number of serious methodological shortcomings.

Chapter 4 reviews representative research on order effects because four of the eight experiments reported in PART II of this thesis take information order as their main theme. Three of the experiments investigate information order with respect to hindsight judgments. The fourth investigates order of information and order of witnesses with respect to testimony in a rape trial. The effects upon juror's perceptions of guilt and innocence are the focus of attention here.

This latter experiment, was run and included in this thesis for four main reasons. Firstly, it is a logical progression to go from looking at order effects within judgment under uncertainty to looking at them on their own. Secondly, most experiments which have investigated order effects in juror judgment have presented an inadequate characterisation of courtroom proceedings to mock jurors. This is discussed in detail in Section 3.3. of Chapter 4. Thirdly, if order of testimony or witness is found to affect judgments then such results might have important implications for the legal system. Fourthly, and lastly, if order effects are detected an explanation of the cognitive processes involved would draw on similar material as those invoked to explain hindsight bias.

Part II of this thesis reports eight experiments designed to investigate the effect of outcome knowledge and information order upon social judgment. The first experiment, reported in Chapter 5, looks at how hindsight and foresight judgments are affected by the method of information presentation and order. The second experiment, Chapter 6, not investigating hindsight bias directly, looks at the time of giving outcome knowledge and information order. Chapter 7 reports the third experiment which looks formally at information order and hindsight judgments. Chapter 8, the fourth experiment, investigates order of testimony and witness in a rape trial with respect to juror judgment. The fifth experiment, Chapter 9, makes use of a contemporary social event to look at judgments in foresight and hindsight. The sixth and seventh experiments, Chapters 10 and 11 respectively look at judgments in foresight and hindsight when personal involvement is high. In the case of Experiment 6 concerning the outcome of a pregnancy test; in Experiment 7 Labour and Conservative subjects assessed possible outcomes to the May 1979 General Election. The eighth and last experiment, Chapter 12, used factual materials about sex role differences in occupations to investigate foresight/hindsight judgments.

Part III, the third and final section of this thesis, consists of two chapters. The first, Chapter 13, provides a general discussion to the experiments, focusing on the effect of outcome knowledge and information order upon judgment under uncertainty. Chapter 14 considers the cognitive and inferential processes that may account for the empirical findings. The chapter concludes with a discussion of the general implications for the model of man as an intuitive scientist and suggestions for future research.

PART I

SOCIAL INFERENCE: RATIONAL MAN, HINDSIGHT AND INFORMATION ORDER

CHAPTER 2.

SOCIAL PSYCHOLOGY'S VIEW OF RATIONAL MAN

2.1. INTRODUCTION

The aim of this chapter is two-fold, firstly, to review some representative research from cognitive social psychology and judgment under uncertainty which is providing an alternative paradigm to the rational model of man; secondly, to show how hindsight bias is not an isolated phenomena but one that fits into a fairly coherent body of research. The chapter begins with a general characterisation of a rational model of man, discusses ways in which it is inappropriate when compared with actual performance, and then provides a general characterisation of man as an intuitive scientist. The major part of this chapter, section 2.2, reviews research, including hindsight bias, which has contributed to this view of man. Finally, the chapter considers the nature of bias and error in social inference.

Social psychological research into human judgment, decision-making and attribution is providing a strong challenge to the deep seated and traditionally held view that man operates rationally in a complex and uncertain environment. The following quote from Shakespeare provides a good characterisation of this traditional view:

> "What a noble piece of work is man. How noble in reason, how infinite in faculties, in form and moving, how express and admirable, in action how like an angel, in apprehension how like a god, the beauty of the world, the paragon of animals." (Hamlet, Act II, Scene 2).

The image of man as noble in reason and intuitive in faculties is developed and expanded upon by economic theory. Here "economic man" in the course

of being economic, is also rational. Simon (1955) outlines the essential features as follows:

> "This man is assumed to have knowledge of the relevant aspects of the environment which, if not absolutely complete, is at least impressively clear and voluminous. He is assumed to have a well organised and stable system of preferences, and a skill in computation that enables him to calculate, for the alternative courses of action that are available to him, which of these will permit him to reach the highest attainable point on his preference scale" (p99).

Steinbruner (1974), in equating rationality with an analytic paradigm depicts rational man as follows:

> "The quintessential analytic decision-maker is one who strains toward as complete an understanding as possible of the causal forces which determine outcomes. He seeks to predict the flow of events and, where he has a leverage to manipulate them to his advantage" (p35).

and

> "The major postulate of analytic theories of decision holds that decisions will be taken which maximise value (utility) given the constraints of the situation" (p28).

In brief, rational man is regarded as an unbiased processor of information, objectively assessing all possible alternatives and using rules of inference which enable him to be successful at these tasks. However, as Abelson (1974) points out, this view, as seen by social psychologists, is too presumptive, prescriptive and preemptive. Abelson raises these criticisms because recent research in human judgment and decision making demonstrates that people's logical and mathematical skills, together with their ability to process information, falls short of the demands of a complex and uncertain environment (for reviews see Dawes, 1976; Fischhoff, 1976; Slovic, Fischhoff & Lichtenstein, 1977; Simon, 1957).

In stark contrast to the rational model, recent research depicts people as reality bound, often unable to process the important features of the environment, let alone consider an event in the context of all possible outcomes (Slovic, 1972). This inability to represent a complex environment in a satisfactory way has given rise to the terms "bounded rationality" (Simon, 1957) and "limited subjective rationality" (Abelson 1974). A number of factors account for this; human attentional and processing capacities are limited (Miller, 1956); the misuse of various heuristics (i.e. informal rules) of thinking (Tversky & Kahneman, 1974); biases and errors in information search and assimilation (Nisbett & Ross, 1980); bias in attribution (Ross, 1977); and biases in prediction (Fischhoff, 1980).

Identifying the tasks of decision-making and social inference has shown that even in relatively simple situations people are called upon to perform many tasks. They must identify and classify quite large amounts of information, retrieve information from memory, provide weights to information according to its reliability and validity, evaluate alternatives in relation to several criteria at once, and finally to decide upon a certain course of action. Accordingly, then, they adopt heuristic strategies to ease the "cognitive strain". Such strategies are seen as being potentially sub-optimal because they fall short of the demands of the analytic paradigm.

The current view of man's ability to operate in a complex and highly uncertain environment differs from that of Simon (1957). His view, based upon the idea of "bounded rationality", depicts man as constructing an over-simplified model of the world but acting rationally towards it.

He says:

> "The first consequence of the principle of bounded rationality is that the intended rationality of the actor requires him to construct a simplified model of the real situation in order to deal with it. He behaves rationally with respect to this model and such behaviour is not even approximately optimal with respect to the real world". (p198).

More recent research, however, depicts man as operating with faulty logic as well as over-simplified representations of the world (Slovic, Fischhoff & Lichtenstein, 1977). The present view, then, depicts people as "lacking the correct programmes for many important judgmental tasks" (Fischhoff, 1976,p424) as well as being subject to limitations in their central information processor.

It must be stressed that error and bias is not always the product of such strategies, they often lead to successful judgments or decisions. What this body of research asserts is that they are often misused or relied upon too heavily, when this occurs systematic shortcomings in social inference arise (Nisbett & Ross, 1980).

The question arises, as to whether a coherent paradigm in the Kuhnian sense (Kuhn, 1962) exists of man as an intuitive scientist. In recent years a number of books have appeared which indicate that a competitor to the rational or analytic paradigm (Steinbruner, 1974) is slowly emerging. Jervis, (1976); Steinbruner,(1974); Nisbett & Ross (1980); Axelrod,(1977); Janis & Mann, (1977); May, (1973) all provide excellent examples of this. For example, Jervis (1976) analyses causes of misperception in international politics by drawing upon much of the psychological theory and research in social psychology which is contributing to this new model of man. Drawing upon cognitive consistency principles, bias in information assimilation, attributional biases he seeks to explain, for

example, why "images" (intentions of another nation) are maintained when rational theories would predict that it is irrational to do so.

Kinder & Weiss (1979) argue that the literature is too disparate and "in no way form a neat well-integrated system of thought about decision-making" (p717).. However, with the publication of Nisbett and Ross (1980) it is probably justifiable to argue that there is more coherence and integration in theory and research than Kinder & Weiss admit. Whilst one would not want to say that a full-blown paradigm of human judgment and decision-making exists, there are strong indications that such a paradigm is emerging. The following section provides a detailed review of research indicating this.

Within this emerging paradigm researchers can be seen as employing two strategies: (1) attempting to provide an accurate description of how people actually perform; and (2) constructing normative models which have not been stated formally and which suggest that the analytic paradigm is inappropriate (e.g. Tversky, 1972; Kahneman & Tversky, *in Press*. Tversky (1972) outlines a model of choice based on elimination by aspects. In this model he views choice alternatives as being composed of a set of characteristics. This model depicts people making choice by first providing weights to each of the characteristics. Then the most heavily weighted characteristic or aspect is considered first in relation to the choice alternatives. Those alternatives which do not have this characteristic are eliminated. The second most heavily weighted characteristic is considered next in relation to the remaining alternatives. Those alternatives which do not possess this aspect are eliminated. This process proceeds until only one alternative is left. This then becomes the alternative upon which the decision is based. Elimination by aspects

is not an analytic model of choice because the procedure does not ensure that the superior alternatives are retained and the inferior ones eliminated. Thus there is no guarantee that such a procedure will not lead to bias and error. The model is normative in the sense that it provides a model of how one ought to go about making decisions. It also specifies a procedure which allows people to rationalise or justify their decision and how they reached it. It is not analytic because the assigning of weights and the selection procedure may not be objectively verifiable.

2.2. MAN AS INTUITIVE SCIENTIST

This section provides a detailed review of some of the research which depicts man as an intuitive scientist. It also locates and relates the hindsight bias in relation to this body of research.

An intuitive scientist uses strategies of judgment and information processing which, although generally successful, often lead to serious bias and error. This is because such strategies have no counterpart in a rational or analytic model of scientific endeavour. In the main, simple intuitive strategies are relied upon too heavily, whilst more formal, appropriate rules are often ignored or underused.

Five areas of research are considered in respect to this: (a) the power of cognitive consistency; (b) information search, assimilation and the revision of opinions; (c) heuristics of thinking; (d) bias in attribution; (e) bias in prediction. These areas are not conceptually distinct -- they overlap considerably -- but provide a useful structure. The review in each of these areas is not exhaustive. This is for two reasons, firstly, it is not necessary for the purpose of this thesis to

do so; the empirical work reported in PART II looks in detail at hindsight bias and order of information. The main purpose of this chapter is to demonstrate that hindsight bias is not an isolated phenonenon but one that fits into a large body of research which shows that people cannot and do not live up to the strictures of rationality. Secondly, such an enterprise would be too large an undertaking in the context of this thesis.

2.2.1. The Power of Cognitive Consistency.

> "By the verdict of innumerable social psychological experiments people prefer to keep their actions, beliefs, attitudes and assorted cognitions mutually consistent. This innocent fact has vast repercussions: it means that we tend to interpret new information to be consistent with previous information."
> (Kinder & Weiss, 1978, p718).

The above quote introduces two important points which have emerged from research on cognitive consistency: (1) Balance Theory (Heider, 1958), Congruity Theory (Osgood & Tannenbaum, 1955) and Cognitive Dissonance Theory (Festinger, 1957) all maintain that man acts in order to keep inconsistencies among beliefs, attitudes or thoughts and behaviour at a minimum; (2) such consistency is often achieved through biased information processing and information search. This section considers these points in relation to a general principle of cognitive consistency and, more importantly, lays the foundations for the subsequent section on information search, assimilation and the revision of opinions.

Consistency among beliefs or between attitudes and behaviour is an important assumption that man holds about what it is to be rational. Reduction of inconsistency to achieve a more "balanced" or "congruous"

state comes about by appropriate adjustment of these cognitions. It is the ways in which this is attained that interest researchers in judgment and decision-making. Rather than following the rules of logic the reasoning processes used are characterised as following rules of "psychologic" (Abelson & Rosenberg, 1958). Irrationality does not follow as an inevitable consequence of these rules, they may often prove reasonable procedures provided that adequate justification can be given According to Jervis (1976):

> "Balanced attitude structures do not reveal irrationality if the cognitive consistency can be explained by the actor's well-grounded beliefs about the consistencies existing in the environment he is perceiving." (p119).

However, because rules of "psycho-logic" provide no guarantee that such perceived consistencies exist in the environment, bias and error often occur. Thus, according to Zajonc (1960):

> "{W}hile the concept of consistency acknowledges man's rationality, observation of the means of its achievement simultaneously unveils his irrationality." (p281).

The theories of Balance, Congruity and Cognitive Dissonance are too well known to warrant any detailed description of them here. Of interest in the context of the present chapter, and especially the next section, is what they have to say about the way consistency is achieved. Consequently, we will only consider the research from this perspective.

Balance Theory (Heider, 1946,1948) suggests that people do not just simplify the environment at random. They make as few changes in their cognitions as possible according to a principle of "minimum effort" (Rosenberg & Abelson, 196); Rosenberg, *et al*, 1960). A similar principle is suggested by Congruity Theory. Osgood & Tannenbaum (1955) state that

"changes in evaluation are always in the direction of increased congruity with the existing frame of reference" (p43). Whilst it may be difficult to determine what the existing frame of reference is, as Eiser (1980) points out, the basic idea that the simplest means of achieving consistency remains. Non-random simplification of the environment to achieve consistency in this manner is the reason why, for example, Kinder & Weiss (1976) discuss information interpretation under the heading of cognitive consistency. This provides added support to the views of Slovic, Fischhoff & Lichtenstein (1977),and Fischhoff (1976) discussed earlier.

Congruity Theory provides a further suggestion as to how consistency is achieved through simplification. It is that assertions or information will be disbelieved and ignored when they conflict with strongly held, divergent attitudes or beliefs (Osgood & Tannenbaum, 1955; Jones and Kohler, 1958). This is the well-known "correction for incredulity" and is similar to Sherif & Hovland's (1961) concept of "latitude of rejection" which has implications for ego-involvement, social judgment and attitude change.

Cognitive Dissonance Theory (Festinger, 1957) indicates two things in the context of the discussion here: (1) post-decisional dissonance is reduced by the seeking of and selective exposure to information which provides evidence for the superiority of the chosen alternative (Brehm, 1956; Ehrlich, 1957); (2) goals are inferred <u>after</u> the action has taken place. This retrospective inference of meaning allows people,according to Schutz,(1967) to provide much better rationalisations for decisions than would be possible before they were made. This follows from the account of consistency being achieved through "maximal simplicity",

discussed earlier. Furthermore, as Pfeffer (1978) emphasises, it is similar to some aspects of what Kelley (1971) has to say about attribution theory, as attributions are made after the behaviour has occurred. Such retrospective rationalisation also bears a strong resemblance to Janis & Mann's (1977) concept of "bolstering", in which they characterise decision-makers as engaging in to avoid pre-decisional dissonance. This is discussed in detail in the next section because cognitive dissonance theory only recognises post-decisional dissonance. In order to discuss strategies of reducing pre-decisional dissonance we first need to establish it exists. It is more appropriate to do this later.

In summary, the family of consistency theories (Balance, Congruity and Dissonance) all indicate man to dislike, and seek to reduce, inconsistency among cognitions. This is often achieved by biased information processing and information search. The next section examines the "power of cognitive consistency" in relation to this.

2.2.2. Information Search, Assimilation and the Revision of Opinions.

The power of cognitive consistency to organise our cognitions implies that the way in which information is searched for, assimilated to pre-existing "theories" and opinions revised will also come under its influence. This is so because people can be seen to adopt strategies which will enable them to avoid inconsistency in the first place. The model of man as an intuitive scientist implies that people hold "theories" about the social world. Such theories can be at a high level (constituting collections of beliefs and attitudes) or at a low level (constituting single beliefs or attitudes). In either case these preconceptions or theories will determine the way information is processed and assimilated. This section

reviews some of the empirical evidence supporting these claims.

The concept of assimilation, used in the following discussions, bears a strong resemblance to Piaget's (1929) use of the term. For him new information is assimilated in such a way as to change the environment. Only when that information is "accommodated" does any change in the self or the person's "schemas" take place.

The claim that information is assimilated to fit with prescribing theories is not a new one, Bacon in 1620, for example, says the following:

> "The human understanding when it has once adopted an opinion draws all things else to support it and agree with it. And although there be a greater number and weight of instances to be found on the other side, yet these it either neglects and despises, or else by some distinction sets aside and rejects, in order that by this great and pernicious predetermination the authority of its former conclusion may remain inviable."

That people hold on to opinions in the face of equally valid, conflicting evidence is demonstrated by the primacy effect in impression formation (Asch, 1946; Luchins, 1957; Hovland, Janis & Kelley, 1953). Professional scientists themselves are accused, by some philosophers of science, of holding on to theories in the face of overwhelming disconfirming evidence (Kuhn, 1962; Lakatos, 1970; Feyerabend, 1975). If professional scientists exhibit such a bias then it is hardly surprising that research by social psychologists reports the same phemomena in the intuitive scientist.

Research into self-fulfilling prophecies (Jones,1977) demonstrates that expectations about ourselves and others influences both our own behaviour and that of others in such a way as to confirm these expectations (Shaklee & Fischhoff, 1979). Self-fulfilling prophecies operate in two

ways: (1) where people see what they expect to. Chapman & Chapman (1969) show people perceive expected relationships in information when it is not there; (2) where an agent causes an actor to behave uncharacteristically but in a way expected by the agent (Snyder and Swann, 1978a,b; Snyder, *in Press*). Snyder & Swann (1978a), for example, show that a perceiver's falsely held beliefs about another person cause him to behave in a manner likely to confirm those beliefs. In testing out one's beliefs, Snyder & Swann (1978a) argue "reality testing has become reality constructing" (p159). According to Snyder (*in Press*):

> "Confirmatory hypothesis-testing strategies may constrain targets to behave in ways that provide actual behavioural evidence that will appear to confirm the hypothesis under scrutiny" (p25).

Labelling theory (Scheff, 1974) and the effects of stigmatisation (Goffman, 1963) provide good examples of this.

Cognitive dissonance theory predicts that defective information search or biased exposure to consistency-maintaining types of information in order to bolster the chosen alternatives is the major way of reducing post-decisional dissonance (Festinger, 1964). What also appears to occur is that biased information processing takes place before a decision is made. This happens because people experience pre-decisional dissonance as well. Empirical support for this comes from a number of studies, for example, Janis & Mann (1977); Fellner & Marshall (1970); Simmons *et al* (1973). Janis & Mann (1977) show,on the basis of numerous interviews with decision-makers, that people employ numerous "bolstering" tactics when appraising possible choice alternatives. One such tactic they discuss is that of

"exaggerating favourable consequences" another is "minimising unfavourable alternatives" (p91). In the case of the former decision-makers are found to pay most attention to the favourable consequences of a choice alternative in order to justify the costs and risks involved. Such attention in itself may well make the favourable consequences seem more likely, as Carroll (1978) has shown. In the case of the latter tactic the decision-maker is found to characterise the undesirable outcomes to a preferred alternative as less bad than they really are, again allowing him to rationalise the costs and risks involved. To summarise, where people have preferences for certain alternatives it is often found that pre-decisional dissonance will arise as other alternatives may be objectively equally attractive. In order to reduce this dissonance "bolstering" strategies may be employed, this means that information search or assimilation may often be biased.

However, information assimilation to pre-existing opinions, beliefs or theories is not something that necessarily leads to bias. As Jervis (1976) points out "intelligent decision-making in any sphere is impossible unless significant amounts of information are assimilated to pre-existing beliefs" (p145). However, circumstances arise in which people maintain their theories or beliefs in the face of clearly disconfirming or contradictory evidence (Schum, 1977; Engquist *et al*, 1976 ; Tucker & Rowe,1979; Erlhom & Hogarth, 1978; Wyer, 1977; Ross *et al*, 1975; Lord *et al*, 1979; Massad, Hubbard & Newson, 1979).

This position is summarised by Kinder & Weiss (1978), who say that "The pursuit of consistency becomes irrational as persistence and denial come to dominate openness and flexibility" (p710). Inappropriate perseverence of beliefs or theories is demonstrated in two studies by

Ross, Lepper & Hubbard (1979) using a "total discrediting" or "debriefing" paradigm. Subjects were given the task of sorting suicide notes, distinguishing those which were authentic from those which were not. They were given one of three types of false feedback: (a) success -- indicating that they had made much more accurate judgments than the average student; (b) failure -- indicating that they had made much worse judgments than the average student; or (c) average -- indicating that they had been as successful in correct categorisation as the average student. Following this subjects were "debriefed" and told that the feedback they had received had been completely false. Ross, Lepper & Hubbard explained at length to the subjects that the feedback they had received had been pre-arranged and determined randomly. They even showed subjects the actual schedule which determined which of the above three conditions the subject had been assigned to and the pre-determined feedback given. In short, subjects were told that the way they had classified the notes bore no relation to how the experimenter had informed them that they had classified the notes.

The results for both participants and observers revealed an extensive degree of post-debriefing perseverence. When asked how many answers (a) they thought they had got correct; (b) they would get correct on a second trial; (c) the average student would get correct; it was found that subjects in the success condition estimated a significantly greater number correct than subjects in the failure condition. Subjects in the average condition fell in between. Furthermore, subjects rated their own ability at such a task as greater when they experienced "success" rather than "failure". Observers, who watched actors perform the task

also showed perseverence in a similar way. In a further attempt to reduce perseverence, the second experiment included a "process-debriefing condition". Here subjects received the same debriefing information as outlined above together with a description and explanation of perseverence itself. It was found that participants demonstrated little perseverence, but actors still showed the effect.

Further demonstration of perseverence is given by Ross & Anderson (1979). They and Ross, Lepper & Hubbard (1979) argue that perseverence reflects the operation of two processes in social-judgment. Firstly, that initial impressions, once formed, are difficult to change -- this relates to the primary effect in impression formation (see Chapter 4 for more detailed discussion). Secondly, and more importantly, they propose that "first impressions may not only be enhanced by subsequent biases in coding but may ultimately be sustained through such biases" (p889). In general then, once an initial scheme for assimilating and coding data is arrived at people are extremely reluctant to abandon it and formulate another.

The idea of a scheme is similar to Jervis' (1976) notion of an "image". He argues that one of the most important tasks for decision-makers formulating foreign policy is to discover the intentions of the other nation. Jervis shows that 'images' of others' intentions determine how information is processed and that images are often retained, irrationally, in the face of overwhelming conflicting evidence.

Maintenance of beliefs or "theories" by biased assimilation of new information is demonstrated in a study by Lord, Ross & Lepper (1979). They used the issue of capital punishment for murder and selected subjects who either strongly opposed or strongly favoured it. Both groups were shown two "real" studies, one supporting and one opposing its deterrent

effect, in a counterbalanced design. There were three main findings: (1) the study which supported subjects' position was rated as more convincing and better conducted than the conflicting study; (2) if the first study subjects read supported their position it served to strengthen it; however, no change in opinions was found if the first study read opposed their position; (3) subjects were found to be more convinced of their view after reading both studies than before reading any of them.

The first two findings do not necessarily imply that irrational procedure for processing information is being employed. As Lord, Ross and Lepper (1979) point out: "our subjects' main inferential shortcoming ...did not lie in their willingness to process evidence in a biased manner. Willingness to interpret new evidence in the light of past knowledge and experience is essential for any organism to make sense of and respond adaptively to, its environment" (p2107).

However, the third finding does constitute serious bias and a departure from rationality. To quote Lord, Ross & Lepper (1979), subjects' "sin lay in the readiness to use evidence processed in a biased manner to bolster the very theory or belief that "initially justified" the processing bias"(p2107). It is tempting to say that the subjects used a form of "grooved thinking",(Steinbruner 1974), and what Jervis (1976) calls "excessive and premature cognitive closure" (p187).

Jervis argues in agreement with other research (for example, Bruner 1957; Haber, 1966), that assimilation of information to pre-existing beliefs or theories is obviously necessary if decision-makers are to function at all in such a complex environment. However, it becomes irrational (i.e. excessive and premature cognitive closure occurs) when

such beliefs have become inappropriate. This occurs when relevant information is ignored or discredited or when the rules for assimilating information exhibit bias. The two experiments outlined above demonstrate just such bias in assimilating and processing information.

In summary, Ross, Lepper & Hubbard (1975) and Lord, Ross & Lepper (1979) demonstrate that not only do beliefs, impressions, etc., persevere in the face of information which discredits the validity for holding them, but also that presentation of mixed evidence causes beliefs, impressions, etc., to be strengthened. Mahoney (1977) shows similar processes operating with professional scientists. He reports that journal article reveiwers are more critical of experiments which conflict with their own theoretical perspective than they are of experiments which confirm their own views.

In conclusion, the research reviewed in this section greatly extends and supplements the way in which the striving for cognitive consistency affects our judgments and decision-making processes. Perseverence and biased assimilation of information demonstrates that initially formed opinions and strongly held beliefs are maintained, often quite inappropriately, by the power of congitive consistency. In many cases our cognitive apparatus makes it easier for us to reduce complexity and conflicting information by assimilating it to pre-existing beliefs, which we may have little or no grounds for holding. Such a strategy is often adaptive and correct, but at other times is the source of serious error and bias.

2.2.3. Heuristics of Thinking.

The ways in which we make judgments under uncertainty in a complex social environment does not appear to accord with rational models of man (Simon, 1957; Slovic, 1972; Abelson, 1974; Steinbruner, 1974; Nisbett and Ross, 1980). Rather than use formal rules of inference which, if

correctly implemented, would guarantee success man is characterised as operating with a number of informal rules or heuristics of thinking (Kahneman & Tversky, 1972, 1973, 1979; Tversky & Kahneman, 1971, 1973, 1974; Ajzen, 1977). These heuristics which are used to reduce potentially complex inferential tasks to more manageable, simpler judgmental operations, often lead to serious shortcomings and bias. Of course, they are often used with success. However, because they have no analogue in a formal model overuse and inappropriate use lead to undesirable consequences. This section describes some of these heuristics and indicates how they may lead to bias and error. Links with other biases and errors discusssed in other sections of this chapter are also indicated. The heuristics of thinking to be discussed provide explanations of the other shortcomings in social inference. It is beyond the scope of this chapter to provide any great detail to the latter consideration -- the interested reader is referred to Nisbett and Ross (1980); Tversky, Kahneman & Slovic, (1980) where greater discussion is given to this.

In making judgments, inferences and decisions, Tversky & Kahneman propose that people make use of simple intuitive heuristics rather than using sophisticated, formal rules. Three heuristics -- availability, representativeness and anchoring and adjustment -- are examples of this. These are not explicit rules of thumb that people use but are "automatic and non-reflective and notably free of any conscious consideration of appropriateness (Nisbett & Ross, 1980, p18).

(a) The Availability Heuristic.

The availability heuristic is used by people when engaged in the task of estimating the frequency of some category or objects or the

probability of some event occurring. Simply stated, the frequency of a class or probability of an event is judged by the ease with which relevant instances or occurrences can be brought to mind (Tversky & Kahneman, 1974). Availability is regarded as an "ecologically valid cue for the judgment of frequency because, in general, frequent events are easier to recall or imagine than infrequent ones" (Tversky & Kahneman, 1973 p209). That is, experience

> "has taught us that instances of large classes are recalled better and faster than instances of less frequent classes, that likely occurrances are easier to imagine than unlikely ones, and that associative connections are strengthened when two events frequently co-occur. Thus a person could estimate the numerosity of a class, the likelihood of an event, or the frequency of co-occurrances by assessing the ease with which the relevant mental operation of retrieval, construction or association can be carried out" (Tversky & Kahneman, p208).

Bias and error occur with the use of the availability heuristic because factors other than frequency or probability (for example, the efficiency of memory search) affect the ease with which instances can be retrieved from memory.

Tversky & Kahneman (1973) report a number of studies which demonstrate how use of the availability heuristic may lead to bias. In one study subjects were presented with the names of famous personalities of both sexes. Some of the names were better known than others. The task was to estimate whether the list contained more males or females and to recall the names on the list. It was found that subjects consistently overestimated the frequency of the sex that had more famous names and recalled more very well known names than less well known ones.

In another experiment subjects were asked to judge the relative frequency of two types of English words -- those beginning with the letter R or those where the third letter is R. Of 152 subjects 105 judged words beginning with R to be more frequent. In fact words with R as the third letter are much more numerous. Bias due to availability occurs here because people find it easier to generate words beginning with the letter R than with R as the third letter. The most available words are those which are generated easiest and hence frequency estimation is biased.

For similar reasons Coombs & Slovic (1978) show that newspaper reporting of more dramatic and exciting causes of death leads people to perceive those causes of death as more frequent than less dramatic, rarely reported ones. In fact, the latter causes are often more frequent than the former. Lichtenstein *et al* (1978) provide additional support for this.

Rothbart *et al* (1978) show that people with more extreme personality characteristics were judged (wrongly) to appear more frequently than those with "mild characteristics" in an impression formation task. They argue that extreme characteristics are more available from memory because of their vividness. This is a similar explanation to that of Coombs and Slovic (1978); Lichtenstein *et al* (1978).

Estimating the probability of occurrance of some social event or outcome can also be influenced by the availability heuristic. This may be due to the ease with which relevant stories, senarios or "scripts" (Schank & Abelson, 1977) can be brought to mind. How plausible such scenarios are, the number of relevant scripts that can be imagined, or the difficulty in producing scenarios may all affect likelihood assessments through availability.

This is demonstrated by Ross *et al* (1977). They found that subjects who were provided with a causal explanation as to why some event should occur from a specified set of antecedent events gave higher likelihood assessments to that outcome than those not provided with a causal explanation. Carroll (1978) has gone even further and shown that just asking subjects to imagine an event turning out in a certain way increases its perceived likelihood. Taylor & Fiske (1975, 1978) generalised this and showed that any strategy which focuses people's attention on potential causes affects their judgment.

Future expectations can also be biased by the availability of instances from past experience. Slovic, Kunreuther & White (1974) show that protection against natural hazard, such as flooding, is often inadequate because people are unable to imagine how future floods could be any different from ones they have experienced. Thus precautions are taken which would be successful with past floods but not necessarily able to cope with future ones. Slovic & Fischhoff (1979) make a similar point with respect to hazard management in nuclear energy.

In summary, availability appears to be a widely used inferential tool. Its misuse and inappropriate application leads to bias. In general, the availability heuristic is most likely to be used when individual instances, scenarios, etc., are used to judge the frequency or probability of events. When generic features come to the fore the representative heuristic is most likely to be employed. It is to this that we now turn.

(b) The Representativeness Heuristic.

The representativeness heuristic (Kahneman & Tversky, 1972,1973; Tversky & Kahneman, 1974) is used to judge the probability of an event or a sample "by the degree to which it is: (i) similar in essential properties to its parent population; and (ii) reflects the salient features of the process by which it is generated" (Tversky & Kahneman, 1972,p431). This heuristic permits people to simplify inferential judgments to ones of simple similarity. An event or instance will be assigned to one category rather than another depending upon the degree to which its salient features represent that category rather than another. In relying on this heuristic, people tend to pay greater attention to more salient, vivid and redundant features of an event rather than base their judgments on more diagnostic information such as base rates and the reliability and validity of the data.

Kahneman & Tversky (1972) report a number of studies demonstrating how the inappropriate use of representativeness leads to biased subjective probability estimates. In one study, for example, subjects were shown the following sequences of coin tosses:
(i) H T T H T H (ii) H H H H T H and (iii) H H H T T T.
They were then asked to estimate the likelihood of each. The three sequences are equally likely, however, subjects chose the sequence which they thought best represented a random process of generation. In consequence, (i) was perceived as more likely than either (ii) or (iii). People appear to expect the essential characteristics of a process of chance to be represented in small samples. The "gamblers' fallacy" -- where deviations occurring in one direction are expected to be balanced by deviations in the opposite direction -- is another example of the representativeness heuristic being misapplied. Deviations are not

corrected, but diluted. In another study, Kahneman & Tversky show how the use of representativeness leads people to ignore prior probability of occurrances or base-rate information when making judgments. Here subjects were provided with short personality descriptions of a number of people who, they were told, had been drawn at random from a group of 100 engineers and lawyers. In one condition subjects were told that the group consisted of 70 engineers and 30 lawyers. The subjects' task was to assess the probability that a personality description belonged to an engineer or a lawyer. It was found that probability estimates reflected more their stereotypes of lawyers and engineers and how representative the description was of these rather than any attention being paid to base-rate information. Misuse of representativeness also accounts for poeple's misconception of regression (i.e. that output should be representative of input). Furthermore, Kahneman & Tversky (1973) have shown that people's confidence in predictions increases as redundancy increases. Redundant information is treated as diagnostic because it is perceived as representative.

In general, then, when people are engaged in predictive tasks requiring them to make subjective probability assessments the representativeness heuristic is greatly overused. More relevant and diagnostic base-rate information is often ignored. Recent research in attribution theory based on Kelley's (1967) covariation model has shown that people often ignore or underuse consensus information when making causal attributions (Nisbett & Borgida, 1975; Nisbett, Borgida,Crandall & Reed, 1976). Consensus information in attribution theory is equivalent to base-rate information in prediction and judgment tasks.

Base-rate or consensus information is not totally ignored however. Certain conditions have been found to encourage its use. For example, when (a) it is given a causal interpretation (Ajzen, 1977; Wells & Harvey, 1977); (b) social desirability for the information is high (Zuckerman, 1978); (c) the base-rate information is about the self (Hansen & Donoghue, 1977); (d) making discrete predictions rather than subjective probability estimates (Manis *et al*, 1980). Base-rate information is also found to be used by asking subjects to make repeated judgments (Fischhoff, Slovic & Lichtenstein, 1979), and by making information concrete rather than abstract (Carroll & Siegler, 1977).

In summary, representativeness is used by the intuitive scientist to assess the likelihood of some event given information of some other state of affairs. Judgments made in such a manner relate features of similarity between specific instances and generalisations. Perceptions of similarity often lead to bias because of the inappropriate use of representativeness. Discussion of the criteria by which people relate the representativeness of one event to another is beyond the scope of this chapter. Tversky (1977), Bar-Hillel (*in Press*) provide some important insights into this.

(c) The Anchoring and Adjustment Heuristic.

The anchoring and adjustment heuristic (Tversky & Kahneman, 1974) is where a decision to start with some "baseline" or natural starting point is made and explicitly revised in the light of new information. The original decision or "anchor" is one that the person may make himself or one that is suggested to him by somebody else. The revision or "adjustment" is often inadequate, as Tversky & Kahneman (1974) show,

because usually it does not take sufficient account of the new information. This results in people being biased towards their original decision or "anchor".

Tversky & Kahneman (1974) demonstrated this, by giving subjects an intuitive numerical estimation task, where, in 5 seconds, one group had to estimate the product of 8 x 7 x 6 x 5 x 4 x 3 x 2 x 1, and another group was asked to estimate the product of 1 x 2 x 3 x 4 x 5 x 6 x 7 x 8. The median estimate for the latter group was 512, for the former 2,250 (the correct answer is 40,320).

The phenomena of "conservatism", found within the Bayesian approach to opinion revision (see Slovic & Lichtenstein, 1971 for a review) is a result of the use of the anchoring and adjustment heuristic. Conservatism is where people, upon the receipt of new information, make smaller changes in estimates of probability than Bayes' Theorem would prescribe, although the changes are in the right direction (Peterson and Beach, 1967; Phillips & Edwards, 1966). Subjects do not make as much use of this new information as this normative standard would prescribe. This is an example of anchoring and adjustment because we can regard the prior probability as the anchor and the revision brought about by the new information as the adjustment. Conservatism, then, results from insufficient adjustment.

In terms of social inference and judgment the anchoring and adjustment heuristic may be generalised from numerical prediction partly to explain the primary effect in impression formation (Asch, 1946) and perseverence in attribution (Ross, Lepper & Hubbard, 1979). In both these cases the anchor is the initial impression or opinion formed, new information does not cause sufficient change as prescribed by strictly logical standards.

Anchoring, in the sense discussed above, is an important psychological concept because it is strongly related to theory maintenance and the power of cognitive consistency, as discussed in section 2.2.2. and 2.2.1. respectively. For example, maintaining a theory in the face of overwhelming conflicting evidence may be little more than a strategy aimed at reducing cognitive strain. An "anchor" or prior theory has this benefit because it provides some framework for processing and searching for new information. In such cases the anchor may and probably did have had adequate justification at the time. Bias enters because new information may indicate this to be unjustified. It is cognitively less demanding to make minor modifications to the original view than to make a radical reappraisal and formulate an entirely different view.

Anchoring and adjustment, then, like the other two heuristics is a strategy people use for reducing cognitive strain in tasks which potentially demand complex information processing. As a result of this and because it is an informal rule of inference it often leads to bias and error on the part of the intuitive scientist.

(d) The Causal Heuristic.

Ajzen (1977) proposes that, in addition to the above heuristics of thinking people hold intuitive theories of events and that the judgmental strategy allowing people to introduce these theories to explain events is a causal heurisitc. This means that in making predictions people give primary consideration to those factors which they believe could cause the event or behaviour to occur. Tversky and Kahneman (1977) argue this also. It follows that information

or data which has no obvious causal significance for people will tend to be ignored.

This, Ajzen (1977) argues, is precisely what happens in experiments which report base-rate information to be ignored when making predictions or judgments. To demonstrate that people use a causal heuristic he conducted three experiments in which base-rate information was given causal significance. In one of these experiments subjects had to estimate the probability that Gary W. had passed a certain examination. In addition to subjects receiving a personality description of Gary W. they also received either "causal" or "non-causal" base-rate information. The causal base-rate information informed subjects that about 75% of students who took the exam had passed (or failed). The non-causal information had no direct implications as to how easy (or hard) the exam was. Here subjects were informed that an educational psychologist had selected a sample from those who took the exam in which 75% passed (or failed). Ajzen found that the causal base-rate information had a stronger and more correct effect on prediction of success (or failure) than non-causal base-rate information.

The intuitive psychologist's assumption that all events, social or otherwise, can be given a causal explanation may lead to severe bias as it is inappropriate in some circumstances. Consider Tversky and Kahneman's (1971) finding that people have inaccurate perceptions of the laws of chance. They found people to believe that a small sample drawn randomly from a population should be representative of the population itself. Hence, failure to find some significant effect

in a replication experiment using a sample size of only half that of the original study may be expected. However, as Tversky & Kahneman (1971) point out, the believer in the "law of small numbers" rarely attributes a deviation from expectation to variability. Rather, he finds a "causal explanation for any discrepancy" (p109). The causal heuristic, then, has such power that discrepancies which should be expected on statistical grounds are actually seized upon and give an inappropriate, erroneous causal explanation.

(e) Hypothesis Confirmation.

The heuristics of thinking discussed so far -- availability, representativeness, anchoring and adjustment, and causality are supplemented by other strategies the intuitive scientist uses when searching for and retrieving information. In addition to the earlier discussion of this, in section 2.2.2., which derives from the power of cognitive consistency, people can also be seen to use an "hypothesis confirmation" strategy. This is in contrast to a disconfirmation approach to hypothesis testing (Wason & Johnson-Laird, 1965; Snyder and Swann, 1978a,b; Snyder, *in Press*). In addition a bias towards memory for confirmatory evidence is demonstrated in a study by Snyder and Uranowitz (1978).

Here, subjects first read a detailed case history of a woman and were subsequently told that she was either a lesbian or a heterosexual Probing of information memorised by subjects showed that those who were told she was a lesbian remembered those aspects of her behaviour that were consistent with such a life style. In a similar way, subjects who believed the woman to be heterosexual remembered information consistent with that life style. Furthermore, incorrectly remembered material was

biased towards behaviour consistent with the life style subjects believed the woman to follow. Information search, as well as retrieval, has also been shown to be biased in a similar way. Snyder & Cantor (1979) showed that information search proceeds by a strategy of hypothesis comfirmation rather than disconfirmation. These studies supplement and extend the discussion of self-fulfilling prophecies given earlier.

The use of the heuristics of thinking together with a confirmation strategy in judging the probability or frequency of an event will mean that the most available information will not only be influenced by such things as perceptual salience, vividness, concreteness, but also that positive instances will be more available from memory. Both because they will be more numerous and regarded as more diagnostic. Further, in imagining an outcome people will tend to imagine how it could come about rather than how it could not. The confirmatory bias, by the same token, will influence the representativeness heuristic in a similar way.

(e) Summary.

This section has discussed some of the informal rules of thinking that people use in decision-making and judgment under uncertainty. It was shown that although they often lead to successful ways of reducing complexity and organising the environment they often lead to serious bias. Often the heuristics are overused or misapplied mainly because they have no analogue in a formal model. These heuristics, chiefly availability, and representativeness, are used to explain how some of the other biases and inferential shortcomings discussed in this

chapter came about. Indeed, Nisbett & Wilson (1977) go so far as to say that our own explanations of the way in which environmental stimuli affect us are inaccurate. They say that these explanations reflect culturally or experimentally derived rules rather than actual cause-effect explanations. They contend that such "explanations" depend, to a large extent, upon availability and representativeness.

2.2.4. BIAS IN ATTRIBUTION.

In general terms attribution theory takes as its central concern the attempt to characterise how man understands and gives causal explanation to his own and others' behaviour. It is a model of man as an intuitive scientist which seeks to characterise his intuitive perceptions of causality.

Heider (1958) provided the initial impetus, and Jones & Davis (1965); Kelley (1967, 1972) developed the ideas in more formal and greater detail. Kelley's model is perhaps the most general for studying attributional processes; that of Jones & Davis deals almost entirely with internal attributions, and focuses on determining the circumstances which promote internal (dispositional), or external (situational) attributions.

The fundamental principle within Kelley's model is that of co-variation. This states that the intuitive scientist examines the presence or absence of three different types of information to determine the type of attribution made. The three types of information are: (1) consistency -- does the actor behave in the same way at different times? (2) distinctiveness -- does the actor behave in the same way in many different situations? (3) consensus -- do other actors perform the same behaviour in similar situations? How these factors co-vary, Kelley claims,

determines whether internal or external attributions are made. High consistency, low distinctiveness and consensus make it more likely that internal attributions will be made. McArthur (1972) has provided the most elaborate test of the model and largely confirms the hypotheses derived.

Fischhoff (1976),in an important paper, has argued that "{A}ttribution researchers find people to be effective processors of information, who organise their world in a systematic manner prone to relatively few biases" (p419). More recent research however (e.g. Ross, 1977;1978) has concentrated on describing the ways in which peoples' perceptions of causality are biased and depart from normative criteria. This section briefly reviews four of these attributional biases -- the fundamental attribution error; actor/observer differences; false consensus and self-serving biases. Finally, some reasons for these two opposing images of man derived from attribution research are discussed.

(a) The Fundamental Attribution Error.

The intuitive scientist suffers from a pervasive shortcoming called the fundamental attribution error (Ross, 1977). This is the "tendency to underestimate the importance of situational determinants and overestimate the degree to which actions and outcomes reflect the actor's dispositions" (p193-194). An experimental demonstration of this is provided by Ross, Amabile & Steinmetz (1977). In this study subjects were either assigned the role of questioner or answerer. The questioner was asked to make up a number of questions from his own general knowledge, he then gave them to the other subject to

answer.' This procedure obviously puts the latter at a strong situational disadvantage, as the questioner can draw upon his own idiosyncratic store of general knowledge. Questioners, contestants and subsequent observers of the 'quiz' were asked to rate the general knowledge of the questioner and contestant. It was found that the contestants and observers all rated the questioner as having superior general knowledge to the contestants. This was also true for the questioners, but less marked. These results show that perceivers ignored the situational determinants which gave such power to the questioner and laid too much emphasis on dispositional aspects of the questioner when making ratings.

Further evidence of the fundamental attribution error is provided by Bierbrauer (1979). In this study observers watched a re-enactment of Milgram's (1963) famous experiment, and then filled in a questionnaire designed to assess the amount of situational and dispositional attributions observers gave to the subject in the experimental enactment. It was found that observers failed to perceive the strong situational factors involved. They gave reasons for the 'teacher' administering high shock levels in dispositional terms, thus holding him personally responsible.

The fundamental attribution error is found in both the intuitive psychologist and his professional counterpart. For example, Mischel (1979) argues that the failure to find cross-situational consistency by personality theorists and the unwarranted adherence to the primacy of dispositions can be seen as due to the fundamental attribution error. Mischel (1968) argued along similar lines but without the

benefit of this research. Jervis (1976) argues that in forming and maintaining images of other nations the policy-maker is apt to consider less the situational forces at work in interpreting their behaviour and more likely to attribute the behaviour to dispositions.

An explanation of why the fundamental attribution error should occur can be traced to at least two sources. Firstly, the tendency to ignore base-rate or consensus information when making predictions or attributions. This is so because such information concerns environmental or situational factors. Secondly, misuse of representativeness and availability. For example, a single piece of behaviour is viewed as representative of the behaviour the person normally engages in, and hence results in dispositional attributions.

(b) Actor/Observer Differences and the Fundamental Attribution Error.

According to attribution researchers, people asked to explain their own behaviour often do so in situational terms. Conversely, when asked to explain the behaviour of another person there is a tendency to attribute stable dispositions (Jones, Rock, Shaver, Goethals & Ward,1968; Jones & Nisbett, 1971; Nisbett, Caputo, Legant & Maracek, 1973). Jones and Nisbett (1971 characterise this actor/observer difference as follows: "there is a pervasive tendency for actors to attribute their actions to situational requirements, whereas observers tend to attribute the same actions to stable personal dispositions" (p80). This divergence in the perceptions of actors and observers is a special case of the fundamental attribution error, with actors being less susceptible to it than observers.

Actor/observer differences may be due to (a) bias, (b) observers having a less complete picture than the actor does, (c) the failure to make a distinction between reasons and causes (Buss, 1978; Locke & Pennington, 1980). Bias may arise when information is made salient because it is vivid or more available from memory. On the other hand actors have "privileged access" to information which observers do not have at all. Hence it is not obvious that this divergence between actors and observers is a bias.

Jones (1976) in considering the above points in more detail, describes the underlying processes that may account for these divergent perceptions. For example, actors are more aware and in touch with the exact situational forces that are working upon them. Observers, on the other hand, regard the actors' behaviour against the environmental background. Thus "different aspects of the available information are salient for actors and observers and this differential salience affects the course and outcome of the attribution process" (Jones & Nisbett, 1971, p85). This relates back to the "vividness" criteria discussed earlier in this chapter. To the extent that there is a difference in the information processed by actor and observer and that this difference is due to salience and vividness.the divergent perceptions may indicate bias. However, the actor does have "privileged access" to information that the observer does not. For example, the actor may know that he behaved in a similar way in similar past situations; if the observer does not know the actor very well or has never encountered him in similar

situations, then he has to make attributions without this knowledge. This may encourage the observer to make dispositional attributions.

(c) Self-Serving Biases.

Attributions made by actors following success or failure at some task has attracted attention from attribution researchers. Research here has been directed at attempting to discover if actors attribute success to personal characteristics and failure to situational forces. Following from this the question of whether motivational or ability variables should be held responsible if internal attributions are made is investigated.

Initial research seemed to show that attributions are made for self-serving purposes (Streufert & Streufert, 1969; Miller, 1976; Bradley, 1978; Miller 1976). In contrast, Ross *et al*,(1974) and Miller and Ross (1975) argue that there is little empirical support for self-serving biases. Miller & Ross concede that there is some evidence consistent with the self-enhancement proposition (subjects take the credit for success) but explain it in cognitive rather than motivational terms. In their view "self-enhancing" effects arise because people (a) tend to expect their own behaviour to produce success; (b) discern a closer covariation between behaviour and outcomes in the case of constant failure; (c) misconstrue the meaning of contingency (p213). However, others disagree (e.g. Weary, 1979) and the debate is still going on.

Amidst the confusion of this area of research in attribution two things emerge: firstly, empirical support for self-serving biases is not unequivocal; secondly, whilst motivational explanations may be possible, there are good reasons for regarding the bias, when it occurs,

as a product of faulty information processing and oversimplification.

(d) The "False Consensus" Bias.

Earlier in this chapter it was shown that base-rate information or consensus information tends to be underused or ignored in certain kinds of social inference, unless, for example, it is made causally relevant. Ross (1977) has argued that, in other kinds of social inference, "laymen tend to perceive a "false consensus", that is, to see their own behavioural choices and judgments as relatively common and appropriate to existing circumstances while viewing alternative responses as uncommon, deviant and inappropriate" (p352). Accordingly, "the intuitive psychologist judges those responses that differ from his own to be more revealing of the actors' stable dispositions than those responses which are similar to his own" (*ibid.* p352).

Ross, Greene & House (1977) report a series of experiments which demonstrate "false consensus" or "egocentric bias". In one experiment subjects were presented with four hypothetical choices (for example, whether or not they would allow comments they had made about a supermarket to be used in a TV commercial). Each subject had to estimate how many of their peers they thought would agree or disagree; indicate what they themselves would do; and rate the traits of a typical individual who they thought would adopt each of the two choices. In accordance with the false consensus hypothesis it was found that subjects who chose a particular alternative estimated that response as more likely among their peers than the alternative choice. Subjects gave less extreme ratings to the "typical" person who they thought would make

the same choice as them. They made more extreme ratings to the "typical" person who would make the alternative choice.

In another experiment, subjects were confronted with a situation that entailed real consequences rather than hypothetical ones. In this case the wearing of a sandwich board advertising a restaurant and walking round campus for 30 minutes with it on. Similar questions were asked of subjects as in the previous experiment. It was again found that perceptions of consensus among peers was strongly determined by their own behavioural choice. Trait ratings followed a similar pattern also.

Goethals, Allison & Frost (1979); Goethals & Darley (1977) provide further empirical support for "false consensus". For example, Goethals, Allison & Frost (1979) report three studies which demonstrate that people holding certain opinions believe a larger and more diverse group shared the same opinion than held different opinions. In one experiment, subjects asked to rate President Carter's performance as good, fair or poor overestimated the proportion of others that agreed with them. Furthermore, subjects perceived people whose opinions were the same as their own to be more diverse than those whose opinions differed from their own.

This false consensus effect, as Ross has argued (Ross, 1977; Ross, Greene & House, 1977) provides a partial explanation of why actor/observer differences should occur. If the observer sees the actor's response as different to that he would make, the actor's behaviour may be labelled as "deviant" and treated as more diagnostic of the actor's

dispositions than behaviour similar to his own.

One of the reasons why the false consensus effect occurs is that "it results from non-random sampling and retrieval of evidence and from idiosyncratic interpretation of situation factors and forces" (Ross, Greene & House, 1977, p299). Here the authors are arguing that biased information search, assimilation and retrieval occurs because people selectively expose themselves to certain groups of people -- often those that behave similarly to us. In doing so the most available and representative information that comes to mind when making judgments about opinions and behaviour is biased. The available instances and representative behaviours are mostly in agreement and consistent with our own.

In summary, the false consensus bias is a powerful source of error in social judgment and inference. It results from inadequate information search proceeding along the lines outlined in section 2.2.2. of this chapter.

(d) Summary.

Having looked at the biases that occur in attribution we now return to the question raised at the beginning of this section. It will be recalled that Fischhoff (1976), when reviewing the research, summarised the findings to date as follows: "{A}ttribution researchers find people to be effective processors of information who organise their world in a systematic manner prone to relatively few biases" (p419). From our discussion of the fundamental attribution error, actor/observer differences, self-serving biases and the false consensus bias, this

would appear not to be the case. In the light of this evidence we might want to say that people are inefficient gatherers and processors of information who organise their world in a systematically biased way. Why is there this discrepancy in research findings?

One answer is that there has been a shift away from the early normative approach of Kelley's (1967) analysis of variance cube towards description of actual attributional processes (Nisbett and Ross , 1980). Another, perhaps more substantial, reason concerns man's ability to detect co-variation. The early research (e.g. McArthur, 1972) presented information to subjects which was relevant and unbiased, and with co-variations that were specified in advance. Thus the task of estimating co-variation between possible causes and effects did not arise. Also the early research does not look at the way in which the intuitive scientist goes about collecting, assimilating and retrieving information from memory and is not called upon to specify the type of information relevant to the problem. Once these aspects become the focus of attention with respect to estimating co-variation, then, as earlier sections of this chapter have shown, bias and systematic error occur at every stage.

The classic phenomena of "illusory correlation" (Chapman & Chapman 1967, 1969) shows that when people assess co-variation -- here inferring psychopathological personalities -- they use pre-existing hypotheses and theories in making judgments. Furthermore, Smedslund (1963) shows that the concept of correlation in the intuitive scientist relies almost exclusively upon the detection of the frequency of present-present instancies in judging relationships (this is a similar strategy to

would appear not to be the case. In the light of this evidence we might want to say that people are inefficient gatherers and processors of information who organise their world in a systematically biased way. Why is there this discrepancy in research findings?

One answer is that there has been a shift away from the early normative approach of Kelley's (1967) analysis of variance cube towards description of actual attributional processes (Nisbett and Ross , 1980). Another, perhaps more substantial, reason concerns man's ability to detect co-variation. The early research (e.g. McArthur, 1972) presented information to subjects which was relevant and unbiased, and with co-variations that were specified in advance. Thus the task of estimating co-variation between possible causes and effects did not arise. Also the early research does not look at the way in which the intuitive scientist goes about collecting, assimilating and retrieving information from memory and is not called upon to specify the type of information relevant to the problem. Once these aspects become the focus of attention with respect to estimating co-variation, then, as earlier sections of this chapter have shown, bias and systematic error occur at every stage.

The classic phenomena of "illusory correlation" (Chapman & Chapman 1967, 1969) shows that when people assess co-variation -- here inferring psychopathological personalities -- they use pre-existing hypotheses and theories in making judgments. Furthermore, Smedslund (1963) shows that the concept of correlation in the intuitive scientist relies almost exclusively upon the detection of the frequency of present-present instancies in judging relationships (this is a similar strategy to

would appear not to be the case. In the light of this evidence we might want to say that people are inefficient gatherers and processors of information who organise their world in a systematically biased way. Why is there this discrepancy in research findings?

One answer is that there has been a shift away from the early normative approach of Kelley's (1967) analysis of variance cube towards description of actual attributional processes (Nisbett and Ross, 1980). Another, perhaps more substantial, reason concerns man's ability to detect co-variation. The early research (e.g. McArthur, 1972) presented information to subjects which was relevant and unbiased, and with co-variations that were specified in advance. Thus the task of estimating co-variation between possible causes and effects did not arise. Also the early research does not look at the way in which the intuitive scientist goes about collecting, assimilating and retrieving information from memory and is not called upon to specify the type of information relevant to the problem. Once these aspects become the focus of attention with respect to estimating co-variation, then, as earlier sections of this chapter have shown, bias and systematic error occur at every stage.

The classic phenomena of "illusory correlation" (Chapman & Chapman 1967, 1969) shows that when people assess co-variation -- here inferring psychopathological personalities -- they use pre-existing hypotheses and theories in making judgments. Furthermore, Smedslund (1963) shows that the concept of correlation in the intuitive scientist relies almost exclusively upon the detection of the frequency of present-present instancies in judging relationships (this is a similar strategy to

hypothesis confirmation discussed earlier in this chapter). Hence, when attribution researchers do not specify co-variation in advance it is hardly surprising to find them subject to certain biases.

In summary, recent research reports extensive and pervasive sources of bias in attribution processes. This contrast with early findings has come about because researchers are now focusing their attention on describing the way in which the intuitive scientist collects, assimilates and retrieves information when making estimates of co-variation. Such considerations as the power of cognitive consistency and the use of various heuristics of thinking also contribute to an explanation of various attributional biases, most notably the fundamental attribution error and the false consensus bias.

2.2.5. BIAS IN PREDICTION

Perhaps one of the most important tasks for the intuitive scientist operating in an uncertain environment is his ability to predict what is likely to happen in the future. As our environment becomes increasingly complex (Steinbruner, 1974) greater demands are going to be placed on our cognitive processes. We have already seen in previous sections of this chapter that the cognitive strategies employed by the intuitive scientist fail to take account of the degree of complexity existing in the world. It was found, for example, that predictions of social events often ignore base-rate information, conform to the "law of small numbers", misperceives causality (Chapman and Chapman, 1969), overuses or misapplies certain heuristics of

thinking (Kahneman & Tversky, *in Press*). This section considers a number of other biases in prediction.

Before describing these other biases an important distinction between "process" and "effect" needs to be made. Cognitive processes, such as the heuristics of thinking, may have the result, through misuse of producing biases in prediction. Hindsight bias, which will be discussed in the next section, may itself be explained by "availability" and "representativeness." However, the relationship between processes and effects is not unidirectional. For example, hindsight bias may play an important role in determining the availability of information on a future occasion. With the increasing amount of research into the model of man as an intuitive scientist this will become a focal point in the future. It is beyond the scope of the present chapter to give any more detailed discussion to this distinction. The final chapter of this thesis (Chapter 14) takes up this distinction.

(a) Hindsight Bias.

In predicting what we know or would have known we may cast our minds back to some past event and attempt to determine how accurately or likely we would have been able to foresee such an outcome obtaining. Our ability to do such a task and the confidence we feel about our accuracy is important because it provides us with strong guidelines concerning what to expect in the future.

As discussed in Chapter 1, historians and those who make use of historical material view events from a privileged position -- they know how things have turned out (hindsight). We intuitively feel that

hindsight confers us with some wisdom and, consequently, enables us to make a more accurate analysis of the event. This view has not been borne out by recent research. Fischhoff and his colleagues (Fischhoff, 1975a,b; Fischhoff & Beyth, 1975; Fischhoff, 1977; Slovic and Fischhoff, 1977) and other researchers (Wood, 1978) have experimentally investigated how possesion of outcome knowledge (hindsight) affects people's perceptions and subjective likelihood assessments of events. They report that being told how an event has turned out leads judges to perceive the outcome they believe has occurred as more likely than had they not known the outcome (foresight). This has been called hindsight bias or the "knew-it-all-along" effect. This bias is reported to demonstrate that people are unable to ignore outcome knowledge when attempting to reconstruct a foresight state. Furthermore, this bias is regarded by Fischhoff (1975a) as seriously curtailing our ability to learn from the past, as indicated by the quote from Fischhoff (1975a) given in Chapter 1. In failing to realise how uncertain and surprising the past was, we may be less capable in assessing the likelihood of future events, and will, in turn, continue to be surprised by how they turn out.

Possession of outcome knowledge, and the increased subjective likelihood assessments associated with it does not constitute bias *per se*. An important conceptual distinction between "creeping determinism" and hindsight bias or the "knew-it-all-along" effect needs to be made. Fischhoff (1975a) hints at this distinction but fails to draw it out fully, and often confuses it in other papers (e.g. Fischhoff, 1980). As will be shown, creeping determinism does not necessarily constitute bias, it may be entirely natural. However, the "knew-it-all-along" effect is a bias.

Creeping determinism refers to the increased probability or likelihood that judges in hindsight assign to outcomes which they believe have occurred. It is a phenomena which occurs when judges are provided with outcome knowledge and only asked to assess the likelihood of that outcome (and others) in the light of the information provided. Judges here are not requested to ignore outcome knowledge or to attempt to reconstruct a foresight state. Hindsight bias, on the other hand is where increased subjective probabilities are found for the outcome believed to have occurred (and decreased probabilities for those believed not to have occurred) when judges are instructed, in some form, to ignore the fact that they know what has happened. This applies whether judges are explicitly asked to ignore outcome knowledge when making subjective likelihood assessments or asked to reconstruct a foresight state using less explicit instructions.

Contrasting the two, creeping determinism, may or may not be justified or rational, as we shall see. The "knew-it-all-along" effect, on the other hand, cannot be justified and is a true bias. Hindsight bias only exists if a person, in trying to reconstruct a foresight state, fails to adequately take account of the privileged view that outcome knowledge provides.

Many events occur in which we are entirely justified in updating probability estimates in the light of outcome knowledge. For example, consider a research chemist who is trying to produce a new kind of chemical. Suppose that theory indicates that by combining two other chemicals in a certain way the desired chemical will be produced.

Before the experiment is carried out the chemist may only feel 65% certain that the reaction will go as predicted. The chemist carries out the experiment and obtains the desired chemical. As chemical reactions are subsumed under a determinist philosophy, the chemist would be justified, in hindsight, in saying that combination of the two chemicals in that way is virtually certain in the future to produce the desired chemical. If, on the other hand, we ask our chemist to cast his mind back to before the experiment was carried out and ask him to reassess the likelihood of the reaction going the way he predicted an estimate greater than 65% would be evidence for hindsight bias. In short, when creeping determinism goes hand in hand with philosophical determinism the former may well be rational or justified.

In the case of social events or assessing one's own or others' factual knowledge, however, it is much more likely to be the case that creeping determinism is imposed upon rather than legitimately inferred from the available information. One reason for this concerns the unique nature of social events. No two events, however similar, can be said to be exactly the same. In summary, an important distinction between creeping determinsim and hindsight bias needs to be made. The latter is an inferential bias, the former may also be a bias but needs further considerations to demonstrate this is the case.

Having made this distinction we will now go on to examine some of the consequences and ramifications of hindsight bias. Fischhoff (1975) has shown that outcome knowledge exerts a strong influence over which antecedent items of information are perceived as relevant to describing

the event. This is an important finding, and one which requires further research, because information considered relevant in this way may be more easily assimilated. This would have the consequence of making it more readily available from memory on later occasions, so providing a complimentary explanation to the account given to biased information search and assimilation in section 2.2.2. of this chapter.

Two other consequences follow from the way in which hindsight biases perceptions of relevant information. First, in searching for information to explain why an event turned out as it did only information confirming the outcome might be selected. Second, selective processing of relevant information might well serve to fuel the hindsight bias. If such information is more available from memory in the future causal explanations as to why it should have come about may be more easily constructed. The more plausible or greater number of causal explanations seem even more inevitable. Ross *et al* (1977); Carroll (1978), provide empirical support for this conclusion.

Hindsight bias can also be shown to influence the way in which future events are predicted. This can be seen most clearly by considering the two ways in which foreign policy-makers misuse history, according to May (1973). Firstly, in dealing with current events policy-makers are over-influenced by the way in which past events have turned out. Secondly, they oversimplify a current international crisis and relate it to the first past analogy that comes to mind without seriously considering the appropriateness of the analogy. For example,

May argues that intervention in Korea in 1950 occurred because "the President and his advisers perceived a North Korean attack on South Korea as analogous to instances of Japanese, Italian and German aggression prior to World War II" (p52). Once an analogy has been made, May further argues, events and facts are seen as conforming to this image. We have already discussed earlier in this chapter some of the cognitive processes that explain why this should happen. The point to be made here is that hindsight bias has a crucial role to play also.

In seizing upon an inappropriate or ill-considered analogy from the past policy-makers are too well aware of how it turned out. If, in hindsight, the outcome is seen as more inevitable than it actually was in foresight, applying the analogy to current events will make a similar outcome, if the same measures are used, appear more likely than it actually is. Thus measures thought appropriate to avoid events turning out as they did in the past may be inappropriately applied to the current crisis. Consideration of policy making in international affairs provides dramatic illustration of the consequences of hindsight bias. However, there is no reason why the same or similar principles should not apply to domestic affairs and events in our own personal lives.

In summary, hindsight bias is a potentially serious inferential error on the part of the intuitive scientist. It may appear at all levels of social judgment, decision-making and prediction and lead to serious miscalculations, misperceptions and biased information processing. Chapter 1 of this thesis has demonstrated generally and

by further illustrated example, the potential importance and significance of this phenomena, little more will be said about it here. Chapter 3 provides a detailed review and critical evaluation of the experimental evidence for it. Further description of the methodology and findings is left until then. A discussion and explanation of the cognitive processes involved is left to PART III of the thesis. Here it need only be mentioned that the heuristics of thinking of Tversky & Kahneman, information assimilation and retrieval, and attribution biases play a central role in this discussion.

(b) Other Biases in Prediction.

Further potential sources of bias in prediction stem from overconfidence in judgments (Lichtenstein, Fischhoff & Phillips, 1977); poor calibration (Lichtenstein & Fischhoff, 1977); value as a determiner of subjective probability (Slovic, 1966) and preference for concrete rather than abstract information when making predictions (Borgida & Nisbett, 1977).

Overconfidence in predictions is found when a person is not only requested to give some quantitative estimate but also to indicate the degree of confidence he has in this estimate. Recent research has reported a consistent bias in confidence estimates. People have been found to be overconfident in evaluating how correct their own general knowledge is (Koriat, *in Press*; Fischhoff, Slovic and Lichtenstein, 1977; Fischhoff & Slovic, 1976). For example, Fischhoff, Slovic & Lichtenstein (1977) asked subjects to estimate the likelihood that certain factual statements were true (for example, that absinthe

is a precious stone). They found that subjects who felt certain that a statement was true were often quite wrong.

The appropriateness of subjective probabilities is determined by their degree of calibration (Lichtenstein & Fischhoff, 1977; Lichtenstein, Fischhoff & Phillips, 1976). A well-calibrated judge is one who is correct, for example, on about half of the occasions that he assigns a subjective probability of 0.5 or odds of 1:1. Similarly, he is well calibrated if he is correct about three-quarters of the time he assigns a subjective probability of 0.75 or odds of 3:1 and so on. Lichtenstein & Fischhoff (1977) report that judges seem to be quite well calibrated when odds are low (i.e. between 1:1 and 3:1) but become less well-calibrated as the odds increase. For example, they find that subjects who gave odds of 1×10^6:1 or more were correct only 90% of the time. Thus instead of giving such high odds they should have been giving odds of around 9:1. Furthermore, Fischhoff & Slovic (1978) indicate that this over-confidence is difficult to counteract.

Over-confidence of this sort does have distinct psychological advantages -- it is one way in which we can organise and feel a certain amount of control over our environment. That it is often misplaced may lead to serious error because an outcome or consequence of a decision we think highly likely may form the basis of a decision or policy. That it is less likely and a greater degree of uncertainty exists means that it may too often go wrong.

One of the possible explanations for overconfidence may be that the value of an outcome or event may be important in determining our

subjective probability of its occurrance. In general, bias in prediction may enter because certain states of affairs are desired. This idea has a long history, for example, McGregor (1938); Cantril (1938) both report that wishes and attitudes, among other things, are important determiners in the prediction of social events. More recent research has also found similar effects (Pruitt & Hoge, 1965; Slovic, 1963; Sevon, 1975; Milburn, 1978, Crozier, 1979). Milburn, (1978) for example, shows that in the prediction of future events positive, and hence highly valued, events were seen as more likely over time. Negative events were seen as less likely over time. For the four time periods 1980-1989; 1990-1999; 2000-2009; 2010-2019, subjects had to assess the likelihood of possible social events such as hunger and poverty no longer being a problem in the United States.

In addition to overconfidence and value the intuitive scientist has a preference for concrete rather than abstract information when attempting to make predictions (Nisbett & Ross, 1980). Earlier in this chapter the research which demonstrated that people ignore base-rate information (even when it would provide more predictive power) was discussed. We now turn to consider how and why concrete or vivid information is preferred when making predictions. Nisbett and Ross (1980) argue that vivid information is more likely to be stored and recalled from memory than is more abstract, statistical data. This may be so because such information has greater emotional interest, provides more detailed and specific knowledge and is more easily assimilated because it is perceived as having high causal relevance. Contributing to this is the idea that actual experience of an event will have greater impact on prediction and judgment than

will, say, second hand reports. This area has received little empirical enquiry from social psychologists to date, most of the evidence for this is anecdotal and derives from case histories (Nisbett & Ross, 1980). From what has been discussed earlier in this chapter it is easy to see why people pay more attention to vivid or concrete information. Such information is more readily available from memory, consequently those engaged in the task of prediction will be more influenced by it. This area should receive more attention from social psychologists in the future. The research will need to demonstrate, firstly, the extent of the preference for this type of information; and secondly, ways in which people can be encouraged to use more abstract data. Ajzen (1977), as previously discussed, has provided preliminary research into this. He demonstrated, it will be remembered, that giving base-rate information a causal interpretation results in it being used to a greater extent, and hence aids people to make more accurate predictions.

In summary, this section has indicated a number of serious shortcomings in the intuitive scientist when he is engaged in the task of prediction. Particularly important in its own right and from the point of view of this thesis is the reported bias of hindsight. Outcome knowledge, it was suggested,not only makes events seem more inevitable but provides a partial account for the biased search and assimilation of information discussed earlier in this chapter. Other biases in prediction discussed in this section may also have a direct bearing on hindsight bias. For example, the most available information in hindsight might well be that which is vivid rather than abstract. If the

abstract information indicates that another outcome might have occurred then it is even more likely to be ignored when making retrospective analyses.

In general then, biases in prediction discussed in this section and referred to in other sections impose serious shortcomings on the intuitive scientist to operate effectively and efficiently in a highly complex and uncertain world.

2.3. BIAS AND ERROR

Throughout this chapter the idea of bias and error in the intuitive scientist has been a common theme. Little or no explicit consideration has been given to these terms; it is important to do so in order to establish more clearly what is meant by them and because it is going to become a major research area in the future (Schneider, Hastorf & Ellsworth, 1979).

Bias, as it has been used in this chapter and in the literature generally, does not refer to individual, idiosyncratic prejudices or occur as a result of certain personality traits. It can be characterised as referring to systematic error which results from faulty or misused cognitive processes that people employ in social inference, judgment, prediction, etc. Systematic cognitive error of this kind may arise for two reasons: firstly, people may possess the correct programmes but cannot use them efficiently because their information processing capacity is severely restricted (Fischhoff, 1976). Secondly, people do not possess the correct programmes to guarantee

that systematic error will not occur. The research reviewed in this chapter is probably best described as resulting from the latter shortcoming. However, it is not merely that "people lack the correct programmes" (Fischhoff, 1976. p424), but that programmes have no analogue in any formal or rational model of inference or judgment. This additional qualification is needed because, as has been emphasised throughout the chapter, people are by and large quite successful. If they were just incorrect programmes they would have been abandoned long ago as they would not be adaptive and would lead to continual misjudgment. In the light of this Fischhoff's remark could, perhaps, be restated as follows: bias and error result from the same processes that are often successful, the programmes people use have no normative point of reference. Inappropriate use or misuse will often produce systematic error.

Bias is often detected because comparisons are made between a normative or rational model and the way in which people actually perform. Ignoring base-rate information, perseverence, hindsight bias, exemplify this. Such a comparison is the source of a potential problem though: in order to assert that systematic deviations occur it is necessary for the normative models to be correct themselves. A recent controversy highlights this problem. It will be recalled that use of the representativeness heuristic often leads to bias as base-rate information may be ignored or paid insufficient attention to. Cohen (1979, 1980) argues that the normative model used to predict how people should make such statistical judgment is based on the well-developed Pascalian theory of probability. Essentially, he is concerned to demonstrate that if a Baconian theory of probability is

people appear to perform in accordance with its principles. ot intended to discuss any of the details of this. For the suffice it to say that 'bias' in the intuitive scientist pendent upon which normative model one uses for comparison.

The discussion of bias as systematic error resulting from propriate use of non-normative 'programmes' and "systematic, sistent and predictable departure of person perception processes outcomes from processes or outcomes prescribed by a particular odel" (Schneider, Hastorf and Ellsworth, 1979, p.226), will receive uch greater attention in the future. The above discussion has served to indicate what we mean when we say that the intuitive scientist exhibits bias. But it is not enough, we need to know, in much greater detail, precisely what is meant by such a claim.

2.4 CONCLUDING REMARKS

[illegible] rovided a review of representative research based upon a mode. of man as an intuitive scientist. The emphasis has been on the ways in which the intuitive scientist exhibits bias and error in social judgment, inference and prediction. The picture that emerges is not a flattering one - at every stage of social inference serious shortcomings have emerged.

The issue raised at the beginning of this chapter can now be addressed: does the work reviewed here contribute an alternative paradigm to the rational one? One line of evidence for accepting

employed people appear to perform in accordance with its principles. It is not intended to discuss any of the details of this. For the moment suffice it to say that 'bias' in the intuitive scientist is dependent upon which normative model one uses for comparison.

The discussion of bias as systematic error resulting from inappropriate use of non-normative 'programmes' and "systematic, consistent and predictable departure of person perception processes or outcomes from processes or outcomes prescribed by a particular model" (Schneider, Hastorf and Ellsworth, 1979, p.226), will receive much greater attention in the future. The above discussion has served to indicate what we mean when we say that the intuitive scientist exhibits bias. But it is not enough, we need to know, in much greater detail, precisely what is meant by such a claim.

2.4 CONCLUDING REMARKS

This chapter has provided a review of representative research based upon a model of man as an intuitive scientist. The emphasis has been on the ways in which the intuitive scientist exhibits bias and error in social judgment, inference and prediction. The picture that emerges is not a flattering one - at every stage of social inference serious shortcomings have emerged.

The issue raised at the beginning of this chapter can now be addressed: does the work reviewed here contribute an alternative paradigm to the rational one? One line of evidence for accepting

such a view comes from the change in emphasis in recent research in attribution. In 1976 Fischhoff characterised research on attribution and judgment under uncertainty as follows:

> "Attribution researchers find people to be effective processors of information who organise their world in a systematic manner prone to relatively few biases. Judgment researchers reveal people to be quite inept at all but the simplest inferential tasks -- and sometimes even at them -- muddling through life by gratuitously allowing for a lot of error" (p419).

Recent research in attribution, especially that by Ross and his colleagues, has dramatically changed this picture. Attribution researchers now find people prone to quite a large number of serious biases. In this sense the two areas of research have now merged and present a similar image of man. The main reason for this development, as indicated earlier in the chapter, is that attribution researchers, like judgment researchers, are describing the way man actually performs.

A further indication of the coherence of the empirical work comes from its application. A number of people from different disciplines are finding this literature rich and fruitful in explaining issues in their own field. This is evidenced by political scientists (e.g. Jervis, 1976; Steinbruner, 1974; Axelrod, 1976) and philosophers of science (e.g. Goldman, 1978) and social psychologists applying it to provide an understanding of negotiation (Morley, 1980).

Finally, it may be the case that a strong competitor to the rational paradigm, outlined at the beginning of this chapter, is not present at the moment. However, the signs are that over the next few years such an alternative will have emerged.

CHAPTER 3

REVIEW OF EXPERIMENTAL STUDIES OF HINDSIGHT/FORESIGHT JUDGMENTS

3.1. INTRODUCTION

Fischhoff and his colleagues were the first to formally investigate the effect of outcome knowledge upon subjective likelihood assessments. They found, as indicated in the previous chapter, that judges who were given outcome knowledge, and thus operating in hindsight, perceived the reported outcomes as more likely than judges who did not know how events turned out, i.e. operating in foresight. Conversely, there was some indication that outcomes reported not to have happened were seen as less likely by hindsight than foresight judges. This latter finding is reported as being less consistent than the former one.

Fischhoff has labelled this perceived inevitability of events "hindsight bias". It will be recalled from the previous chapter that a distinction between hindsight bias and creeping determinism was made (see pp 52). Because of the importance of this distinction for what follows it will be briefly described again here. Hindsight bias can only be said to occur when a judge is told the outcome and asked to ignore this fact or to reconstruct a foresight state when making subjective likelihood assessments. Creeping determinism occurs when a judge is told the outcome and merely asked to produce subjective likelihood assessments in the light of this and/or other relevant

information. In both cases there is found to be greater probabilities assigned to relevant outcomes by hindsight judges than foresight judges.

This chapter presents a detailed review of six papers (Walster, 1967; Fischhoff, 1975a; Fischhoff & Beyth, 1975; Fischhoff, 1977; Slovic & Fischhoff, 1977; Wood, 1978) which empirically investigate this. These papers are divided into three categories according to the kind of material judges are provided with. (1) Experiments which use "laboratory material". Here subjects are provided with descriptions of social events which are either concocted by the experimenter or are true but unfamiliar historical episodes. Three of the six papers fall into this category -- Fischhoff, 1975a; Slovic & Fischhoff, 1977; Walster,1967. (2) Experiments making use of contemporary social or socio-political events -- Fischhoff & Beyth, 1975. (3) Experiments where material of a factual nature is presented to subjects. Here, for example, judges are presented with a word plus its definition or a factual statement and asked to assess the probability of its being true (or false). Two papers -- Fischhoff, 1977; Wood, 1978 -- are of this type.

The above categorisation is adopted for two main reasons: (1) it provides a useful way of conceptualising the experiments and allows useful and critical comparisons to be made. For example, it is found that the strongest evidence for hindsight bias comes from experiments using factual material. (2) The experiments described in PART II of this thesis also conform to this framework.

The detailed review which follows adheres to a similar pattern for each of the papers considered. First, the experiment(s) reported in each is briefly described and the results summarised. Second, a critical evaluation of the methodology and statistical analysis is given. Last, an overall evaluation of the paper is made. The chapter finally assesses the extent and magnitude of hindsight bias.

3.2. EXPERIMENTS USING LABORATORY MATERIAL

3.2.1. Fischhoff (1975a)[1]

This paper reports three experiments which systematically investigate the effect of outcome knowledge upon judgments under uncertainty. The first is concerned with creeping determinism; the second and third with hindsight bias. Common to all three experiments are the event descriptions used (Mrs. Dewar in Therapy; The British-Gurkha Struggle; Near Riot in Atlanta; George in Therapy). Each was a passage of about 150 words with four possible outcomes specified. In the hindsight conditions the "true" outcome was appended to the end of the event description. Subjects made likelihood assessments in terms of percentages and were instructed that the total for the

---

[1] This study is reviewed first because it represents the first systematic attempt to look at judgmental differences in foresight and hindsight. Walster (1967), although chronologically earlier, does not have this as a central issue. Rather Walster's central concern is to investigate how knowledge of the consequences of another's decision affects anticipation of the outcome. Specifically, Walster tests the hypothesis that the more momentous the consequences of a decision the more subjects in hindsight would feel that they could have correctly anticipated it. She also investigates the idea that attribution of responsibility to the decision-maker will be greater when the consequences are more important. In short, her paper does not explicitly look at hindsight/foresight differences.

four outcomes to any event description had to be 100 per cent.

Experiment 1 consisted of one foresight condition and four hindsight conditions (one for each outcome) for each event description. Table 3.1. presents the general design used in this experiment. Using an independent measures design, subjects in all conditions were asked to "evaluate the probability of occurrance

Table 3.1. General Design of Experiment 1 in Fischhoff (1975a).

	Foresight	Hindsight 1	Hindsight 2	Hindsight 3	Hindsight 4
OUTCOME A		"True"			
OUTCOME B			"True"		
OUTCOME C				"True"	
OUTCOME D					"True"

Cells with "True" indicate this to be the outcome hindsight judges were told occurred.

of each of the four possible outcomes in the light of the information appearing in the passage" (p289). The important point to note here is that subjects in the hindsight conditions were not being asked to ignore the fact that they know what has happened or to attempt to reconstruct a foresight state. The experiment, then,focuses on the question of how knowing how an event has turned out influences people's assessments of both that outcome and the outcomes reported not to have happened. Fischhoff found that, in accord with creeping determinism, reporting an outcome increased its percieved inevitability.

Experiment 2 investigated how aware people are of this tendency to see outcomes to events as more inevitable when knowing how they have turned out. Using the same event descriptions as in Experiment 1, but only running hindsight conditions (i.e. 1 - 4 in Table 3.1, not the Foresight condition) and getting one subject to assess outcomes for each of the four event descriptions, Fischhoff provided subjects with the instruction " to respond as they would have had they not known the outcome". That is, subjects are being asked to ignore outcome knowledge when assessing the likelihood of the outcomes to the event description. In analysing the results Fischhoff compared these assessments with those given in Foresight in Experiment 1. He found hindsight assessments to be consistently and significantly higher than corresponding foresight ones. This, he argues, indicates that once people know how an event turned out they are unable to accurately reconstruct a foresight state. In short, they are biased.

Experiment 3 extends this line of enquiry to see whether the hindsight bias is present when subjects are asked to respond as they thought other people would who were not in possession of outcome knowledge. Here subjects were asked to make assessments "as would other students who did not know what happened". In this experiment there was one foresight condition and four hindsight conditions (one for each of the four possible outcomes to an event description). As in Experiment 2, each subject made appropriate judgments for each of the four event descriptions. The results again showed hindsight

estimates for the outcomes believed to have occurred to be consistently higher. Fischhoff argues that this indicates that subjects respond as they believe they would have done and not as others would have had they not known the outcome.

In all three experiments subjects were also asked to make relevance judgments. They assessed, on appropriate seven point scales, "how relevant or important each datum appearing in the event description was in determining the event's outcome (p289). Generally, it was found that outcome knowledge influences perceived relevance. Items which had a direct bearing on a reported outcome were provided as much more relevant and diagnostic than items of information which did not have a direct bearing on the outcome(s) hindsight judges believed had occurred.

Two main criticisms to these experiments can be made, one concerns demand characteristics (Wood, 1978), the other concerns the experimental design and subsequent analysis of the second experiment described above.

(a) Demand Characteristics and Hindsight Bias.

Demand charcteristics[2] (Orne, 1959; 1962; 1967) of the experimental situation is concerned with the way in which subjects perceive

---

[2]Psychological research into the role of the subject in psychological experimentation has been investigated from a number of perspectives. As well as demand characteristics such things as bias in subject selection (Smart, 1966), the volunteer subject (Rosenthal, 1965), experimenter bias (Rosenthal, 1966,1967), and the problem of deceiving subjects (Kelman, 1967) have all been reported to have significant and largely undesirable influences on the performance and responses of subjects.

the research and the way in which they focus on certain unintended cues when deciding how to respond. Orne contends that subjects are likely to "respond in a manner which is likely to support the hypothesis being tested" (Orne, 1972, p779).

A problem of demand characteristics is only appropriate to Experiments 2 and 3 described above. The main problem centres around the fact that subjects, in the hindsight condition, are given a piece of information (here outcome knowledge) and then asked to ignore it when making likelihood assessments and relevance judgments. The question that arises is what subjects make of this "ignore" instruction. The subject might quite justifiably ask himself -- why give this information to me in the first place if I am supposed to ignore it? It might then be realised or reasonably guessed that this is the focus of the experimental hypothesis. Given the propensity, as mentioned earlier, for subjects to want to please the experimenter and produce results in line with his hypothesis it may be the case that the subjects' likelihood assessments, particularly to the outcome believed to have occurred, are so influenced. This would mean that the hindsight bias reported in Experiments 2 and 3 may be interpreted as the result of the demand characteristics of the experiment. Experiment 1 does not suffer from this problem as subjects are only asked to assess the likelihood of a number of outcomes to a given event in the light of <u>all</u> the information given. Where "all" here includes outcome knowledge. It might be the case that undue emphasis is placed on this but as long as creeping determinism

is beihg investigated and not hindsight bias it can hardly be raised as a criticism. Wood (1978) also has a demand characteristic interpretation of hindsight bias uppermost in his mind. He attempts to give subjects instructions to rule this out, but as we shall argue, is not particularly successful.

In summary, a demand characteristic interpretation of hindsight bias is a possibility given the nature of the instructions subjects are provided with. If true it would cast doubt upon the validity and generalisability of the experimental results. One important task, then, is to attempt to provide subjects with instructions which will be less susceptibile to this criticism.

(b) Methodological Considerations.

The main criticism here concerns the design of the second experiment. In contrast to the first one each subject is provided with each of the four event descriptions and asked to assess the likelihood of the relevant outcomes. The second experiment only contains four hindsight conditions (one for each of an event description), no corresponding foresight condition is run here. Two problems arise from this, firstly, the legitimacy of comparing hindsight conditions in Experiment 2 with foresight in Experiment 1. Foresight subjects in Experiment 1 assessed the possible outcomes to only one event description. In contrast, hindsight subjects in Experiment 2 made four sets of assessments, one for each event description. Such things as learning effects, repeated exposure to the same instructions to ignore outcome knowledge may serve to enhance the demand characteristic interpretation. Furthermore, there may have

been differences in the population samples used, Fischhoff uses students from a number of different universities. For the sake of thoroughness in Experiment 2 it would only have required an extra 20 subjects or thereabouts to run a comparable foresight condition. Given the small number of subjects required in relation to the large number used overall it is difficult to understand why such a "control" group was not incorporated into the design. The third experiment does have a foresight condition, of necessity as the instructions differ from Experiments 1 and 2, so the above criticisms do not apply here.

The statistical analysis Fischhoff gives to the data in the three experiments also presents problems. In the summary tables of likelihood assessments in the three experiments means are given. This implies that Fischhoff regards the data as being interval or ratio. Given this it would be appropriate to use parametric statistical tests. However, in testing for differences between hindsight and foresight judgments he uses the sign test and the Median test. These are both non-parametric tests which reduce the data considerably and are less powerful than parametric tests. Given that his sample size is large (around 20 in each cell) he can hardly argue that non-parametric statistics are appropriate on the grounds that he does not have a large enough number of subjects in each cell.

The problem is compound by the way in which subjects' responses are categorised in order to make data appropriate for the sign test.

Fischhoff counts the number of times hindsight assessments are greater than the respective foresight ones for a certain outcome to an event description. If the former is greater than the latter a + sign is allocated, which indicates support of the hypothesis. A sign test is then carried out on the total number of comparisons. This may be warranted in providing an overall picture but it gives no indication of the size of the effect. In a more detailed analysis, which might give this information, Fischhoff uses the Median test. Given that means are presented the most appropriate statistics would be the t-test or ANOVA. The statistical analysis used to look at foresight/hindsight differences in relevance judgments is paradoxical in the light of the above discussion. Relevance judgments, it will be remembered, were made on 7-point scales. With this kind of rating scale it is never clear whether the data should be regarded as ordinal or interval. However, Fischhoff treats it as interval and uses parametric statistical tests.

Overall the statistical analysis has a paradoxical air about it. Where he reports means and regards the data as interval or ration he uses non-parametric statistical tests. Where there is an element of doubt over whether the data is interval or ordinal he uses parametric statistical tests. This is a disturbing analysis by Fischhoff and leads one to speculate why he might have adopted such an approach. For example, in giving means it is usual to give standard deviations also. Perhaps variance was so large that

parametric tests failed to yield significant results? Whatever the reason consistency would require that medians and not means be presented if non-parametric tests were used.

In summary, this review of the three experiments reported in Fischhoff (1975) demonstrates a number of shortcomings in a systematic attempt to investigate differences in foresight/hindsight judgments. The problems of demand characteristics, experimental design and statistical analysis indicate that the reported hindsight bias should be treated with some caution. The first two criticisms only apply to Experiments 2 and 3, the third to all three. Evidence for creeping determinism seems fairly compelling and would be even stronger if we could have more confidence in the statistical analysis. Evidence for hindsight bias meets with greater reservations. In general, before accepting the wider implications of hindsight bias we need to establish that it is more than an artefact and subject the hypothesis to more rigorous and thorough empirical investigation.

3.2.2. Slovic & Fischhoff (1977)

This paper again reports three experiments, which the authors claim "examine the existence and workings of hindsight bias in lay assessments of scientific research" (p544). All three experiments present subjects with short summaries of research reports. Each experiment will be briefly outlined and then critically evaluated.

Experiment 1 used four short (about 150 words) summaries of experiments representative of research in biology, psychology and

meteorology. Foresight subjects were asked to assess the likelihood of two possible outcomes, and explain why each might happen. They were then asked to assess the likelihood that each of the outcomes would be replicated in all, some or none of ten subsequent replication trials of each experiment. Subjects in hindsight were informed of the outcome of the first trial of each experiment and asked to assess the likelihood of replication of the original result in all, some or none of ten subsequent trials of each experiment. Results indicated that those who had been told how the experiment had turned out on one trial (hindsight) thought it more likely that it would turn out in that way on future trials than those who had not been told how one trial of the experiment had turned out (foresight). For six of the eight comparisons there was a significant difference, the other two being in the predicted direction.

Experiment 2 was similar to the first except that foresight subjects assessed the likelihood of only one of the two possible outcomes to each experiment. This was done in order to make hindsight judgments more comparable with foresight ones. Foresight subjects assessed the likelihood of all, some or none of 10 subsequent trials turning out this way <u>if</u> the experiment had turned out in such a way on the first trial. Hindsight subjects followed a similar procedure as in Experiment 1 but were also asked to explain how the study might have worked out the other way. For example, if they had been told that, in the case of the "Gosling Imprinting"

experiment, that the gosling approached the goose, they were asked to explain why it might approach the duck (other outcome). This was asked of subjects in an attempt to reduce hindsight bias. Results obtained were similar to those of the first experiment with the latter instruction having very little effect.

Experiment 3 was designed to counter criticisms concerning the brevity of the reports, hence subjects received full length manuscripts. For the foresight condition the report consisted of introduction and method sections; the hindsight condition consisted of these same two sections together with the results. In both conditions subjects had to assess; (a) the surprisingness of results (in hindsight the assessment concerned the reported result, in foresight it concerned the two possible alternatives); (b) the stability of results (i.e. how likely to be replicated). As predicted it was found that in hindsight subjects found the reported outcomes to be less surprising than foresight subjects. Also, in hindsight the reported outcome was believed to be more replicable than in foresight.

Two problems arise concerning what these experiments are claimed to be testing. Firstly, as quoted earlier, the authors claim to be "examining the existence and workings of hindsight bias" (p544). This is not so, however, due to the central concern being with the extent of replicability of an experiment given that one trial has turned out in a certain way. The problem here is that

hindsight is confounded with people's tendency to generalise or view a small sample as representative of what would be the case in a population or large number of occurrances. This consideration derives from Tversky & Kahneman's (1971) "law of small numbers". Their research demonstrates that people are biased towards seeing a small sample from some population as similar to the population in all important characteristics.

The second consideration which questions the claim that these experiments are investigating hindsight bias centres around the distinction made between hindsight bias and creeping determinism. Hindsight bias only occurs when people are instructed to ignore outcome knowledge when making judgments but are unable to do so (see Chapter 2, pp 52). In the three experiments described here this was not requested of subjects. They were not asked to ignore the result of the first trial when assessing replicability, surprisingness, etc. In view of this the experiments can only, at most, be regarded as examining the existence and workings of creeping determinism. This means, again, as discussed in Chapter 2, that the results found here are not evidence of bias *per se*, it might be the case that the increased probabilities in hindsight are justified. The research needs to establish that they are not for their claim to be investigating hindsight bias to make any sense.

In summary, these experiments do not, as the authors claim, investigate hindsight bias. At most they are looking at creeping determinism. However, it seems much more reasonable to argue that

they are an investigation of Tversky & Kahneman's law of small numbers.

3.2.3. Walster (1967).

The third and final experiment to be reviewed in this section is Walster's (1967) "second guessing important events". The paper reports two experiments, both concerned with examining how confidence in predicting consequences of decisions is affected by the magnitude of the consequences involved. Experiment 1 is concerned with a decision to buy a house, Experiment 2 with taking a job. Walster's main hypothesis was that the more serious the consequences the more outcome knowledge induces the feeling that the consequences should have been anticipated first time around.

Walster described three levels of financial gain (high, medium and low) for both types of decision. Subjects were asked to assess how likely they thought they would have anticipated the consequence they believed had occurred. They were also asked to assess the degree of responsibility they attributed to the person for the consequences. In Experiment 1 a "no information" or control condition was included in which subjects were not told about the financial gain or loss that had occurred as a result of the decision to buy a house. In Experiment 2 two control conditions were included, one similar to that in Experiment 1, and one where subjects were not told whether a decision had been made.

The control conditions in Experiments 1 & 2 can be regarded as foresight conditions; where subjects are informed of consequences is equivalent to a hindsight condition. Results from both experiments demonstrated that the greater consequences (financial gain or financial loss) the more subjects felt they would have been able to anticipate such outcomes themselves. No systematic relationship between the degree of financial loss or gain and attribution of responsibility was found. For control conditions Walster only reports probabilities assigned to various consequences. No direct analysis of "foresight/hindsight" treatments is made.

A number of points can be made about this study in relation to later investigations into differences between foresight/hindsight judgments. First, it does not directly investigate either creeping determinism or hindsight bias. However, if we do look at the experiments from such a perspective we find that the 'hindsight' conditions are similar to those of Experiment 2 in Fischhoff (1975a). The instructions given to subjects by Walster were "suppose you had been told everything about Mrs.W's decision (except, of course, whether she won or lost money as a consequence of her decision). Do you feel you would have been able to guess whether or not the purchase would have had good or bad consequences, just from your knowledge of the situation?" (p241). In effect subjects are being asked to ignore the fact that they know the financial consequences when making confidence assessments, etc. Therefore we can characterise the experiments as implicitly investigating hindsight bias.

Reconceptualised in this way the results show that greater hindsight bias is exhibited when the consequences of a decision are more momentous. This is so because "hindsight" subjects told of a financial gain anticipated less of a loss (the greater the gain the less loss anticipated), and subjects told of a loss anticipated greater loss (the greater the loss the more it was anticipated). This pattern of findings is exhibited in both experiments. These results would seem to suggest, in the absence of proper statistical analysis, that knowing how an event turned out leads people to be biased in that they thought they would have anticipated it in foresight. Because the experiments did not ask subjects to assess the likelihood of each of the six possible consequences direct foresight/hindsight comparisons cannot be made. Accordingly, the results can only be taken as tentative indications of hindsight bias.

3.2.4. Summary.

These three papers (Fischhoff, 1975a; Slovic & Fischhoff, 1977; Walster, 1965) do not lend strong support to there being a bias of hindsight. Fischhoff (1975a). whilst systematically investigating foresight/hindsight judgments, suffers from three major shortcomings: that of a demand characteristic interpretation, inadequate experimental design and statistical analysis. Slovic & Fischhoff (1977), although claiming to be investigating hindsight bias, are, as discussed, looking at creeping determinism at most, and probably the consequences of the "law of small numbers" of Tversky & Kahneman (1971). Walster (1967) obliquely investigates hindsight bias but provides no more than an

indication as to its existence. In all, then, these three studies do not provide strong support for a bias of hindsight.

3.3. EXPERIMENTS USING CONTEMPORARY SOCIAL EVENTS

"Laboratory material", such as has been used in the three papers just reviewed, has provided some indication of the effects of retrospective analysis. However, it is of obvious importance to ensure that these findings do not occur solely in the laboratory. To this end it is, therefore, necessary to demonstrate that hindsight bias occurs in everyday life. Fischhoff & Beyth (1975) report an experiment which made use of contemporary socio-political events.

3.3.1. Fischhoff & Beyth (1975).

The socio-political events used by Fischhoff & Beyth (1975) to investigate hindsight/foresight judgments were President Nixon's 1972 visits to China and the U.S.S.R. In contrast to the three studies just reviewed this one used a repeated measures design. Fifteen possible outcomes were described for each of the visits and subjects assessed the liklihood of each outcome for either the visit to China or the visit to the U.S.S.R. Sometime after the visits had occurred, subjects repeated their tasks, presumably knowing how Nixon's trips had turned out (hindsight). They were instructed to try and remember the probabilities they had given in foresight, or failing that to reconstruct a foresight state and give probabilities

they thought they would have given before the visit had taken place. Some subjects made foresight judgments shortly before the visit and hindsight judgments either shortly afterwards (two weeks) or long afterwards (four months). A further experimental condition was run in which only hindsight judgments were made. This "after-only" group was instructed to "reconstruct probabilities, giving the probabilities which you would have given on the eve of the visit". For each of the fifteen outcomes to each trip subjects, in hindsight, also indicated whether they thought they had happened or not.

The overall picture that emerged from the results was that two thirds of the subjects produced hindsight judgments in the predicted direction. In more detail, three quarters gave higher assessments to outcomes they believed had occurred. A much weaker results was obtained for outcomes believed not to have occurred. Here only fifty seven per cent gave lower assessments in hindsight. It was also found that the bias was more pronounced when hindsight assessments were made long after the visits had taken place than shortly after. In the "after-only" condition hindsight bias was shown by two thirds of subjects for events believed to have occurred, and just over fifty per cent for outcomes believed not to have occurred.

This experiment is much less susceptible to a demand characteristic interpretation of hindsight bias than Fischhoff (1975). This is because asking subjects to cast their minds back to before the outcomes of contemporary events such as President Nixon's visit to China is something we often do in everyday life. The request is not an unusual one.

However, telling people how an event turned out and then asking them to ingore it, as with Fischhoff (1975a)is unusual and unreliable. It is unrealistic in the sense that subjects have no foresight state which they have experienced to attempt to cast their minds back to. With contemporary and experienced events we often try to imagine back to before the result was known in attempting to understand how it came about. By the same token, likelihood assessments made in this way can also be seen as something we often do. In short, the use of contemporary socio-political events within this paradigm has strong attractions due to the confidence we might have in the validity of the findings showing hindsight bias.

Unfortunately, the statistical analysis of the data may be criticised because Fischhoff & Beyth adopt rather loose criteria in determining whether subjects showed bias or not. The procedure used was as follows: each of the subject's responses to the fifteen outcomes was divided up into two categories (i) those subjects believed happened, and (ii) those subjects believed did not happen. For each category foresight (f) and hindsight (h) assessments were compared. If $f < h$ and $f > h$ for (i) and (ii) above respectively a positive (+) sign was allocated, indicating it to be hypothesis supporting. If, on the other hand, $f > h$ and $f < h$ for (i) and (ii) respectively a minus (-) sign was allocated, indicating non-hypothesis supporting. If $f=h$ for either category a (o) was allocated, and the data subsequently dropped from the analysis.

The next step was to sum the number of (+) and (-) for each subject. If there were found to be more (+) than (-) the subject was deemed to be hypothesis supporting, if there were more (-) than (+) the subject was deemed to be non-hypothesis supporting. If the number of (+) and (-) was equal the subject was dropped from the analysis. Finally, an overall index was computed combining these two categories.

Table 3.2., reprinted from Fischhoff & Beyth, shows the number of hypothesis supporting and non-hypothesis supporting subjects for the two categories, together with the overall index, for each of the experimental groups. Also shown in Table 3.2 are Z-values and the percentage hypothesis supporting subjects for each category and experimental group.

The first problem concerns the allocation of a sign (+, -, or o) to each outcome. This method fails to take account of the extent to which hindsight and foresight judgments differ. It may well have been the case that hindsight estimates were only slightly different to foresight ones. If true this method of analysis would overemphasise the strength of hindsight bias.

The second problem concerns the procedure by which subjects are characterised as hypothesis supporting or not. Again the size of the differences are ignored when only the number of (+) and (-) for each category and overall for each subject is computed. This again may tend to overemphasise the extent of the bias. It may have been the case that subjects had only one more (+) than (-).

Table 3.2. Results printed from Fischhoff & Beyth (1975,p8) showing the number of hypothesis-supporting (+) and non-hypothesis-supporting (-) subjects for each experimental group and response category.

DATA FROM FISCHHOFF & BEYTH (1975)			
Group I (N = 29)			
CHINA: Shortly before, shortly after.			
	Believed happened	Believed did not happen	Overall
+	17	15	17
-	7	8	11
%+	70.8	65.2	60.7
Group II (N = 41)			
CHINA: Shortly before, long after.			
+	30	15	26
-	7	22	14
%	81.1	40.5	65.0
Group III (N = 26)			
USSR: Shortly before, shortly after.			
+	14	13	15
-	8	8	8
%+	63.6	61.9	65.2
Group IV (N = 41)			
USSR: Long before, shortly after.			
+	30	15	19
-	5	24	13
%+	85.7	38.5	59.4
Group V (N = 23)			
USSR: Long before, long after.			
+	7	15	18
-	3	6	4
%+	85.0	71.4	81.8

Table 3.2. (continued)

DATA FROM FISCHHOFF & BEYTH (1975)			
Groups I,II,III,V Combined (N = 119).			
Groups I,III,IV,V Combined (N = 119) (In parentheses)			
	Believed happened	Believed did not happen	Overall
+	78 (78)	58 (58)	76 (69)
-	25 (23)	44 (46)	37 (36)
%+	75.7(77.2)	56.9(55.8)	67.2(65.7)
Group VI (N = 27)			
USSR: No before, shortly after.			
+	20	10	16
-	4	14	8
%+	83.3	41.7	66.7
Group VII (N= 27)			
CHINA: No before, long after.			
+	16	13	14
-	5	13	12
%+	76.2	50.0	53.8
Group VIII (N = 37)			
CHINA: No before, long after.			
+	18	22	24
-	13	12	13
%+	58.1	64.7	64.9
Groups VI,VIII Combined (N = 64)			
Groups VII,VIII Combined (N = 64) (In parentheses)			
+	38(34)	32(35)	40(38)
-	17(18)	26(25)	21(25)
%+	69.1(65.4)	55.2(58.3)	65.6(60.3)

Table 3:3. Reanalysis of Fischhoff & Beyth's (1975) data with "draws" included as instances of non-hypothesis-supporting (-) subjects.

DATA REANALYSED TO INCLUDE 'DRAWS'			
Group I (N = 29)			
CHINA: Shortly before, shortly after.			
	Believed happened	Believed did not happen	Overall
+	17	15	17
-	22	14	12
%+	58	51.2	58.6
Group II			
CHINA: Shortly before, long after.			
+	30	15	26
-	11	26	15
%	73.2	36.0	63.4
Group III (N = 26)			
USSR: Shortly before, shortly after.			
+	14	13	15
-	12	13	11
%+	53.8	50.0	57.7
Group IV (N = 41)			
USSR: Long before, shortly after.			
+	30	15	19
-	11	26	22
%+	73.0	36.0	46.3
Group V (N = 23)			
USSR: Long before, long after.			
+	17	15	18
-	6	8	5
%+	73.9	65.2	78.3

Table 3.3. (continued)

DATA REANALYSED TO INCLUDE 'DRAWS'			
Groups I,II,III,V Combined (N = 119)			
Groups I,III,IV,V Combined (N = 119) (In parentheses)			
	Believed happened	Believed did not happen	Overall
+	78(78)	58(58)	76(69)
-	41(41)	61(61)	43(50)
%+	65.5(65.5)	48.7(48.7)	63.9(58.0)
Group VI (N = 27)			
USSR: No before, shortly after.			
+	20	10	16
-	7	17	11
%+	73.9	37.0	59.3
Group VII (N = 27)			
CHINA: No before, long after.			
+	16	13	14
-	11	14	13
%+	59.2	48.1	51.2
Group VIII (N = 37)			
CHINA: No before, long after.			
+	18	22	24
-	21	15	13
%+	48.6	59.4	64.9
Groups VI,VIII Combined (N = 64)			
Groups VII,VIII Combined (N = 64) (In parentheses)			
+	38(34)	32(35)	40(38)
-	26(30)	32(31)	24(26)
%+	59.4(53.1)	(50) (54.7)	62.5(59.4)

The third, and most serious, problem stems from the previous two. It concerns the cases where foresight estimates equal hindsight ones and where subjects have equal numbers of (+) and (-). Fischhoff & Beyth, as outlined earlier, allocate both types of occurrances a (o) and then drop the data/subject in further analysis. This is unfortunate for the following reasons. Subjects, when making hindsight judgments, are explicitly instructed to give "the same probabilities you gave then (two weeks ago). If you cannot remember the probability you assigned then, give the probability that you would have given to each of the various outcomes on the eve of President Nixon's trip...." (p5). Thus accurately remembered or reconstructed probabilities must be regarded as *non-hypothesis supporting, and hence evidence against hindsight bias*. Such instances should be assigned a (-) and included in the analysis. They should not be assigned a (o) and dropped from the analysis.

Strictly speaking, the hypothesis being tested by Fischhoff and Beyth is that "the remembered or reconstructed probability of an event will tend to be larger than the probability originally assigned to it if the event is believed to have occurred, and smaller if it is believed not to have occurred", (p3). Dropping cases if f = h and (+) = (-) gives a misleading picture as to the percentage of hypothesis supporting subjects.

Table 3.3 shows the effects of reanalysing Fischhoff and Beyth's data with cases of foresight = hindsight counting against the hypothesis of hindsight bias, i.e. assigned a minus sign. The figures are obtained by taking an experimental group adding the number of (+) and (-) together and then subtracting this from the total number of subjects (N) in that group. This figure was then added to the non-hypothesis-supporting category. The percentage of hypothesis-supporting (+) subjects was worked out by simply dividing the number in that category by the total N and multiplying by 100.

The percentages shown in Table 3.3 are lower than those in Table 3.2. Whereas 3.2 shows that two-thirds of subjects exhibited hindsight bias Table 3.3 reveals a figure of less than 60 percent overall. For outcomes believed to have happened Fischhoff and Beyth (Table 3.2) report three-quarters of the subjects as giving higher assessments in hindsight. In contrast, Table 3.3 shows this figure to be less than two-thirds. For outcomes believed <u>not</u> to have happened Table 3.2 shows 57 per cent of subjects giving lower assessments in hindsight; the re-analysed data (Table 3.3) shows this figure to be <u>less</u> than 50 per cent. Table 3.3 also shows only three of the eight experimental groups to have greater than 60 per cent hypothesis-supporting subjects for outcomes believed to have occurred. The table also shows only one of the eight experimental groups to exceed this figure for outcomes believed not to have occurred. In contrast, Fischhoff & Beyth (Table 3.2) report seven and four out of eight for outcomes believed to have and not to

have occurred respectively. Re-analysing the data in this way shows the extent and strength of hindsight bias to be considerably less than Fischhoff & Beyth claim.

Unfortunately, inferential statistical analysis cannot be carried out on this re-analysed data in a similar way to Fischhoff and Beyth. The sign test is inappropriate and consequently a Z-score cannot be obtained.

The fourth and final criticism is a general one: although it is desirable to compute some overall index of foresight/hindsight differences it is also necessary to find out how many estimates for each question in the various groups are individually different. This could be achieved simply by comparing individual likelihood assessments using, for example, the Wilcoxon T-test. Such an analysis would provide more detailed information concerning the magnitude of hindsight bias and provide what is lacking from Fischhoff & Beyth's analysis.

3.3.2. Summary.

The problems discussed concerning Fischhoff & Beyth's procedure for analysing their data (a) use a method which provides a very weak criteria as to what is to be considered as a case of hindsight bias; (b) incorrectly analyse the data using this procedure, they ignore what should be regarded as non-hypothesis-supporting instances; (c) give no indication as to the strength or magnitude of the bias. Such shortcomings mean that one should be cautious about the strength of the bias. Re-analysis of the data (Table 3.3) shows the bias to

be less extensive than claimed by Fischhoff & Beyth. In conclusion, stronger empirical support for hindsight bias is required before we can feel confident in its existence outside of the laboratory.

3.4. EXPERIMENTS USING FACTUAL MATERIAL

The two final experiments to be reviewed in this chapter both make use of factual material. As we shall see,Fischhoff (1977); Wood (1978), in contrast to the previous studies, provide the soundest evidence for the existence of hindsight bias, and give some indication to its extent.

3.4.1. Fischhoff (1977).

The two experiments reported in this paper firstly attempt to demonstrate hindsight bias, and secondly, try to discover if the bias can be reduced or eliminated. Both experiments make use of factual material, this was taken from a wide range of areas such as history, music, geography, literature. Subjects were presented with a word or statement and two definitions; they had to indicate which they thought correct. For example, Aladdin's nationality was (a) Persian or (b) Chinese. Subjects had to assign, to one of the alternatives, a subjective assessment of being correct.

Experiment 1 consisted of three treatments. For each treatment subjects in foresight answered seventy five factual questions of the form described above. An hour later subjects in treatment 1 (reliability) were provided with twenty five of the original seventy

five questions and asked to respond as they had done earlier. Subjects in treatment II (memory) were given the same instructions but were also told which was the correct answer to each question. Subjects in treatment III (hypothetical) were given twenty five new questions, told the correct answer, and asked to respond "as they would have had they not known what the answer was" (p350).

It was found that both hindsight treatments (II and III above) induced bias, over seventy per cent of the cases in which subjects did not remember their original response showed this. Subjects in the 'hypothetical' treatment (III) gave higher probabilities for correct answers. This was the case in twenty out of the twenty five test items when compared with foresight probabilities made by subjects in other treatments to the same question. For subjects in both treatments, hindsight bias was less pronounced for incorrect than correct answers. Subjects in the 'reliability' treatment remembered two-thirds of their original responses (compared with only half of subjects in the 'memory' treatment), for those not remembered no systematic effects were found.

Experiment 2 attempted to eliminate or reduce this bias. Using the same material as in Experiment 1, three 'hypothetical' (treatment II above) conditions were run. In the first the same procedure as Experiment 1 was used; in the second (hypothetical debiasing) subjects were told about hindsight bias and asked to do everything they could to avoid it. In the the third (hypothetical warning) subjects were requested to work as hard as possible at the

task, this condition was included in order to counteract the tone of the instructions for the previous group.

Results showed that the new instructions had little effect on hindsight bias. No substantial differences in mean probabilities for the twenty five questions appear for any of the three treatments in this experiment, or with the two hindsight treatments in Experiment 1. This experiment, Fischhoff claims, demonstrates the robustness of hindsight bias.

Two criticisms can be levelled at this study. The first concerns the problem of demand characteristics discussed earlier and applies mainly to the 'hypothetical' group in Experiment 1 and Experiment 2. Here, subjects are being asked "to respond as they would have had they not been told what the answer was". This "ignore" instruction is very similar to that in Fischhoff (1975) and the reader is referred back to the relevant discussion in Section 3.2. of this chapter. Problems of demand characteristics also arise with respect to the 'memory' group in Experiment 1. The one hour time delay between foresight and hindsight judgments with subjects being told the correct answer and asked to remember the probabilities given an hour ago must strike subjects as particularly odd.

In particular it is interesting to speculate what subjects might have thought the experiment was about. For example, being told the outcome and asked to ignore it might lead them to realise

the true purpose of the experiment. Although Fischhoff provides a cover story -- to the effect that it will aid experiments in evaluating the difficulty of the questions -- it is hardly convincing. Subjects might, rightly, argue that such information could be obtained from the assessments given in foresight. In view of this and the discussion given earlier a demand characteristic interpretation of hindsight bias cannot be ruled out.

The second criticism concerns the design of Experiment 2, which did not include a foresight treatment of its own. Instead Fischhoff compared the hindsight ('hypothetical') treatments with the foresight treatment in Experiment 1. It is, however, important to note that both the 'hypothetical debiasing' and 'hypothetical warning' treatments included the following instructions: "Your responses are extremely important to us. The effort you invest in them will largely determine the value of our subsequent study. Please devote as much attention to this task as you can" (p354). As the foresight treatment in Experiment 1 contained no such exhortation to work as hard as possible it is essential to have a foresight group with such an exhortation in Experiment 2.

Overall, this study, especially the first experiment, provides the strongest evidence, from the experiments reviewed in this chapter, for hindsight bias. The second experiment, because of inadequate design, does not clearly demonstrate the robustness of

bias as Fischhoff claims. Of course, the problem of a demand characteristic interpretation of the findings may temper the degree to which the results are accepted.

3.4.2. Wood (1978).

One of the central themes of this study is the problem of a demand characteristic interpretation of hindsight bias. Wood reports two experiments using similar factual material to that of Fischhoff (1977), in which subjects are provided with a cover story to explain why the experiments are as they are. Wood claims this cover story adequately disguises the true purpose of the experiments and consequently reduces the problem of demand characteristics. The experiments also investigate how the bias is affected by amount of outcome knowledge, type of question, and the effect of making foresight and hindsight judgments. In each experiment subjects were provided with a number of factual assertions, half being true and half false. The task was to rate the plausibility of these assertions on a seven-point scale ranging from definitely false to definitely true. Subjects performed this task twice: once in foresight, and once in hindsight. In hindsight subjects were given the correct answers to all questions in one block. Hindsight subjects were given one of two sets of instructions, either 'memory' where they were told that "the primary interest was in their ability to give consistent ratings for those statements presented in Stage 1 and Stage 3". (i.e. foresight and hindsight)(p347). Or, 'peer' where people

were asked to ignore the information (outcome knowledge) given in Stage 2 and rate the statements as they thought other college students would.

The results showed hindsight ratings to be higher for true statements and lower for false statements than foresight ones, in accordance with hindsight bias. No effect of different instructions (peer or memory) was found. Also, a larger bias was exhibited for true statements than false ones, in accordance with Fischhoff (1977).

Expeirment 2, using the same material and similar procedure, investigated the effect of pre-outcome judgments, instructions and the number of times subjects were exposed to the correct answers. The 'memory' and 'peer' instructions were the same as in the first experiment. The main difference between this experiment and the last one concerned the number of exposures to outcome knowledge. One group were shown the true answers and tested only once; another group were shown and tested three times. Additional instructions were also provided in an attempt to disguise the real purpose of the experiment. The instructions given stressed that Wood was interested "in assessing which statements were easy to learn and which were difficult as a way to study memory for the truth value of statements".

In general, Wood found that foresight/hindsight judgments were in the expected directions. Specifically, those exposed and tested on the correct answers three times showed a larger hindsight bias

than those only shown and tested once. Also subjects who made both foresight and hindsight ratings to the same questions showed less bias than those who only made hindsight judgements. This was only true in the case of the 'memory' group, not the 'peer' group.

These results do appear to replicate and extend Fischhoff's (1977) findings. The study indicates that hindsight bias is stronger for true rather than false statements; not affected by instructions to remember original foresight ratings or answer as you think others would who did not know the answer. A repeated measures design finds less bias than an independent one. However, the main question of whether Wood provides adequate instructions to remove fears of a demand characteristic interpretation remains. The 'peer' and 'memory' instructions in both experiments, plus the added 'plausible' instructions given in Experiment 2 are seen by Wood as increasing our confidence that "the effects are real and not an artefact of the experiment procedures used to demonstrate it" (p352).

The 'memory' instructions emphasised consistency, and appear not to ask subjects explicitly to ignore the fact that they know the correct answer. In view of the problems of Fischhoff (1975), discussed earlier, this memory manipulation by Wood would seem to succeed in reducing the likelihood of a demand characteristic interpretation. In contrast, the 'peer' instructions do not achieve this. Here, it will be recalled, subjects are asked to

ignore outcome knowledge and rate as they thought other students would who did not know the correct answer. This represents no improvement over Fischhoff (1977). The additional cover story in Experiment 2, given at the time outcome knowledge is provided does offer some rationale to subjects concerned by why they should make three sets of responses. Once again, however, as with Experiment 1 only the results from the 'memory' group give us some reason to regard a demand characteristic interpretation as unlikely.

3.4.3. Summary

Fischhoff (1977) and Wood (1978) provide the strongest support for the claim that outcome knowledge produces judgmental biases. With the proviso of a demand characteristic interpretation, which Wood goes some way to discounting (though not far enough), the use of factual material within this paradigm would seem to provide us with the best evidence for the existence of hindsight bias.

3.5. OVERALL EVALUATION OF THE SIX STUDIES.

From the detailed, critical review presented in this chapter shortcomings in four areas indicate that the experiments do not demonstrate hindsight bias to be as reliable or as strong an effect as the researchers claim.

One problem encountered was with a demand characteristic interpretation. For Fischhoff (1975a) and Fischhoff (1977) this was shown to cast doubt on the confidence we could have in accepting

their interpretation of the results. Experiments using contemporary socio-political events (Fischhoff & Beyth, 1975) suffered less from this problem. The explicit attempt by Wood (1978) to provide a cover story to subjects succeeded in allaying our fears over this for his 'memory' groups but not for his 'peer' groups. On the whole instructions given to subjects in hindsight to ignore outcome knowledge when making likelihood assessments are particularly open to demand characteristics. To overcome this it is necessary to provide equivalent instructions to the 'ignore' instructions which do not explicitly make this request. This may be a difficult thing to do and such a procedure might be open to criticism that one is merely investigating creeping determinism and not hindsight bias. However, such instructions can be formulated, as will be shown in PART II of this thesis where two experiments have this issue as one of their central concerns.

The second and third problems encountered with the six studies reviewed here were to do with shortcomings in experimental design and statistical analysis. Both Fischhoff (1975a) and Fischhoff (1977) omitted an important foresight condition in one of their experiments. The descriptive and inferential statistics in Fischhoff (1975a) and Fischhoff & Beyth (1975) was shown to make use of inappropriate tests, use weak criteria for establishing when subjects exhibited hindsight bias and provide little indication of the magnitude of the bias.

Fourthly, and finally, Slovic & Fischhoff (1977), it was argued,

is not to be considered as an examination of hindsight bias. At most it could be regarded as looking at creeping determinism. It is more likely that the work is an extension of Tversky and Kahneman (1971). Walster (1967) provides some indications of hindsight bias and its consequences, although her study was not designed with this in mind.

In conclusion, of the six studies here all but Walster (1967) have been shown to suffer from a number of serious shortcomings. Fischhoff (1977) and Wood (1978) provide the best evidence for hindsight bias. The three experiments which make use of "social scenarios", using laboratory or contemporary events, suffer from shortcomings which question the interpretation and the results themselves. In view of these problems it is of obvious importance to establish just how reliable and extensive the hindsight bias is. Furthermore, it is also important to establish the conditions which may encourage the bias and those which reduce or eliminate it. These considerations are of central importance to the experiments to be reported in PART II of this thesis.

CHAPTER 4.

REVIEW OF ORDER EFFECTS AND IMPLICATIONS FOR HINDSIGHT FORESIGHT JUDGMENTS

4.1. INTRODUCTION

At the end of the preceding chapter it was indicated that one of the objectives of the experiments to be reported in PART II of this thesis is to provide some indication of the conditions under which hindsight bias might be produced and conditions which might eliminate or reduce it. One important variable which might be expected to affect subjective likelihood assessments in both foresight and hindsight is the order in which information describing a social event is presented.

Order of information presentation has been found to influence such things as impression formation (Asch, 1946; Luchins, 1962,1965; Manis & Platt, 1979; Zadny & Gerard, 1974), ability attribution (Jones & Goethals, 1972; Jones *et al*, 1972), persuasive communication (Hovland, 1957), and judgments in Bayesian tasks (Peterson & Ducharme, 1967; Shanteau, 1970). These findings may be important if psychology is to be applied to disciplines such as the law (Tapp, 1977). Order of presentation of testimony (Lawson, 1967; Lawson, 1969; Stone, 1969; Costopolous, 1972), defence-prosecution order (Wallace & Wilson,1969), "climax" ordering of arguments (Thibaut & Walker, 1975) have all been reported to have systematic effects on judgments. Given that information order has reliable and predictable effects, as evidenced above, then

by the same token it would seem reasonable to argue that hindsight /foresight judgments might be similarly influenced.

This chapter first reviews research that has investigated order effects in the above areas. No attempt is made to provide a complete treatment, but it is clear, nevertheless, that certain conditions are conducive to producing a "primacy effect", whilst other conditions tend to produce "recency effects". In the review of order effects in psycho-legal studies emphasis is put on Thibaut and Walker's (1975) work in this area. This is for two reasons, firstly, it is widely regarded as an important and substantial contribution to this issue. Secondly, some of the material and procedures they use are drawn upon in the design of a number of the experiments to be reported in PART II of this thesis. The final section of this chapter discusses how the above findings may be applied to investigations into hindsight/foresight judgments. In particular, how order of information might be expected to influence creeping determinism, and, more importantly, hindsight bias.

4.2. GENERAL RESEARCH ON ORDER EFFECTS.

Four main areas will be reviewed here, these being impression formation, ability attribution, Bayesian judgments and persuasive communication. As we shall see from this review there is no simple rule or rules predicting the circumstances for obtaining primacy effects and those for obtaining recency effects. A number of factors have to be taken into consideration, for example, time delay between information presentation and judgments, familiarity with stimulus material.

4.2.1 IMPRESSION FORMATION

Asch (1946) found evidence for a primacy effect in impression formation. Using a set of six adjectives to describe an imaginary person he presented one experimental group with them arranged from favourable to unfavourable (intelligent, industrious, impulsive, critical, stubborn, envious). A second experimental group received the same six adjectives but in reverse order. Substantially different impressions of the imaginary person were formed by the two groups. More favourable impressions were formed when favourable adjectives appeared at the beginning of the list. In contrast, more unfavourable impressions were formed when unfavourable adjectives appeared at the beginning of the list. Similar findings have been reported using different sets of adjectives (for example, Anderson and Hubert, 1963; Anderson and Norman, 1962; Hendrick and Constantini, 1970). However, no primacy effect is found with shorter lists of adjectives (Anderson and Barrios, 1961) and is much less pronounced when subjects are repeatedly exposed to the list (Anderson and Hubert, 1963).

An experiment by Luchins (1957), where two one-paragraph descriptions of an imaginary person were used, also reported primacy effects. Here one of the paragraphs reported the imaginary person as friendly, outgoing and extrovert, the other as shy and introverted. Subjects who received the "friendly, extrovert" portrayal first perceived the person as more friendly than subjects who read the "shy, introvert" portrayal first. However, in another study (Luchins and Luchins, 1970) order of information was not found to have

long-term effects. These studies indicate that first impressions are not long lasting and subject to change with the addition of new information.

We might conclude that it is only when judgments have to be made based purely on first impressions and where information is limited (but not too much so) that information occurring first is going to influence judgments. Under such conditions primacy effects may be explained by change of meaning (Asch, 1946), attention decrement (Anderson & Hubert, 1963) discounting (Anderson and Jacobson, 1965), or due to different belief systems being formed (Fishbein & Ajzen, 1975).

4.2.2. ABILITY ATTRIBUTION.

Attribution theory and attributional biases were discussed in some detail in Chapter 2 so little more will be said about the general area here. Ability attribution, or more specifically, the attribution of ability following success or failure is subject to a strong primacy effect. Jones *et al* (1968) had subjects observing a stimulus person answering a series of intelligence test questions. Three conditions investigated order effects: where the stimulus person (a) began giving a large number of correct answers and then performance declined; (b) began giving a large number of incorrect answers and subsequently improved; (c) performed consistently throughout. Subjects were asked to (i) rate the intelligence of the stimulus person; (ii) estimate how many correct answers the person would give on a subsequent trial; (iii) recall how many correct

answers were given in the original trial. For all three measures it was found that subjects attributed greater ability and intelligence to the person in condition (a) above, in condition (c) the person was rated least able.

These results demonstrate how early appearing information influences predictions of future performance and that it biases what is remembered of past performance. This strong and consistent primacy effect is explained by Jones & Goethals (1972) in terms of people perceiving early information as relatively stable. Subsequent information is then assimilated to this image formed earlier on. Jones *et al* (1972) provide further support for this explanation. This explanation is similar to the general account of theory maintenance and change in Chapter 2.

The studies by Jones and his colleagues have strong implications for judgment under uncertainty generally and hindsight/foresight judgments in particular. This will be dealt with in greater detail later. Briefly, because, for example, Jones *et al* (1968) indicate that both predictions and postdictions are influenced by information appearing first we might expect hindsight judgments to be similarly influenced.

4.2.3. BAYESIAN JUDGMENT.

The Bayesian approach to the study of human judgment and decision-making makes the basic assumption that "opinions should be expressed

in terms of subjective or personal probabilities, and that the optimal revision of such opinions, in the light of relevant new information, should be accomplished via "Bayes' Theorem" (Slovic & Lichtenstein, 1971, p30). Bayes' Theorem is, therefore, a normative model for investigating opinion revision and human judgment. A more detailed discussion of this approach was given in Chapter 2 (Section 2.2.2.): only a general outline will be given here.

Peterson & Beach (1967) indicated that people performed in reasonable accord with the Bayesian model. More recent research, reviewed in detail by Slovic & Lichtenstein (1971) presents a radically different picture. At best people are conservative processors of information, at worst people fail to adhere with Bayesian dictates, their deviations and discrepancies are seen as having no analogue in a normative or formal model.

The occurrence of order effects in tasks where people are required to update subjective probability estimates in the light of new information would be obvious transgressions of Bayes' Theorem. Both primacy and recency effects have been reported. Primacy effects have been found by Peterson & DuCharme, 1967; Roby, 1967; Dale, 1968; recency effects by Pitz & Reinhold, 1968; Shanteau, 1970.

Recency effects using poker chips (Pitz & Reinhold, 1968) and red and white lights (Shanteau, 1970) are found to occur when subjects

are presented with a small number of samples from a population and asked to make subjects probability estimates after each sample. Pitz & Reinhold (1968), for example, had two bags, each containing 100 poker chips. In one bag the distribution was 60 red, 40 blue; in the other 60 blue and 40 red. Choosing one bag at random a trial consisted of 5 draws of one chip with replacement. Subjects who revised estimates after each one of the 5 draws exhibited recency. On the other hand, subjects who made probability estimates only after seeing the result of all 5 draws for any one trial showed no order effects. Using a descriptive, as opposed to normative, approach Shanteau (1970), in testing an additive model of decision-making also reports recency effects. Although it is difficult to compare the results from these two different approaches the similarity of task allows us to state that the conditions outlined earlier are conducive to recency effects.

The conditions under which a primacy effect is obtained within Bayesian and related areas of judgment, and those in which a recency effect is obtained are difficult to specify in a systematic way. Studies which demonstrate primacy all present subjects with long sequences of data (100 in Peterson & DuCharme, 1967), which are structured to point towards a certain outcome. In contrast, studies reporting recency present subjects only with short sequences of data.

The finding that people are, on the whole, "conservative" when revising opinions (Peterson, Schneider & Miller, 1965; Phillips and

Edwards, 1966) is a kind of primacy effect, since subjects fail to update subjective probabilities sufficiently in the light of new information. This can be seen as subjects giving undue weight to information appearing early or to being "anchored" by it with insufficient adjustment subsequently.

In general then, although both primacy and recency effects have been found to occur under most conditions in Bayesian judgment type tasks, one would expect primacy given that the central finding from such approach has been conservatism in judgment (Slovic and Lichtenstein, 1971).

4.2.4. PERSUASIVE COMMUNCIATION.

Perhaps the most sustained research programme to investigate factors which influence peoples attitudes and beliefs and, hence, judgments, is that of Hovland and his colleagues (Hovland *et al*, 1953; Janis & Hovland, 1959). They conceptualise persuasive communication as being composed of three components -- the communicator, the communication, and the respondent. Within this framework they have sought to discover factors to account for successful and unsuccessful communicator or attitude change.

The focus of this chapter on order effects automatically directs our attention to the communication or message component. Research on persuasive communication which has dealt with this has produced the most thorough and detailed analysis of order effects. Further, a systematic account of the conditions most likely to yield primacy

and those most likely to yield recency has emerged. The empirical work is not solely attributable to Hovland's research programme; it was mentioned first for the framework it provides in understanding persuasive communication.

Lund (1925) appears to be the first social psychologist empirically to investigate order effects within this context. The question he investigated was: if both sides of a controversial issue are presented successively, which has the advantage -- the side presented first or last? Lund presented subjects with pros and cons of topical issues such as whether all men should have equal political rights; whether monogamos marriage would continue to be the only socially accepted relation between the sexes. He found that the side presented first had a consistent advantage over the side presented second. He called this the "law of primacy" in persuasion. Later research showed this to be somewhat of an over-statement. According to Hovland (1957): "when two sides of an issue are presented successively by different communicators the side presented first does not necessarily have the advantage".

Subsequent research has shown that a primacy effect may be expected with non-salient material (Rosnow & Goldstein, 1967); where the topic is highly controversial (Lana, 1963a); where subjects express or have a high interest in the topic (Lana,1963b). Recency effects have been found to occur for salient topics (Rosnow and Goldstein, 1967); uninteresting subject matter (Lana, 1963b); and moderately unfamiliar issues (Lana, 1961).

Overriding any primacy or recency effect is strength of arguments for the other side, then the side with the strongest argument has the advantage, regardless of whether he goes first or last (Rosnow *et al*, 1966).

Janis & Hovland (1959) indicate that the order of information or arguments is not the only, or perhaps most important, message factor. One-sided versus two-sided communications differ in their ability to persuade (Lumsdaine & Janis, 1953), as do low-and-high-fear appeals (Higbee, 1969). Furthermore, order of presentation of a communication can be divided into two questions. The first, for which experimental work has been reviewed in this section, is concerned with the most effective order of separate communications. The second concerns what the most effective ordering of the elements in a communication is, i.e. deals with a single source arguing for one side of an issue. The latter question has been less intensively investigated.

The types of issues examined here include such things as the effect of implicit versus explicit conclusions, presenting the conclusion first or last, climax or anticlimax order. Research has shown that drawing an explicit conclusion makes the message more effective (e.g. Hovland & Mandell, 1952; Maier & Maier, 1957). Little empirical work has addressed itself to the question of whether it is more effective to draw a conclusion at the beginning or end of a communication. McGuire (1969) gives a theoretical discussion to this,

arguing that effectiveness depends on whether the information is agreeable or disagreeable to the audience. For the former, giving the conclusion first may be the most effective, for the latter, last may be more effective.

Climax versus anticlimax order arises as an issue when there are a series of arguments in a communication, some of these arguments being stronger than others. A climax order is where the strongest are given last, i.e. "built up to", anticlimax where the strongest are given first. Empirical investigation into this topic appears to be inconclusive, Sponberg (1946) found anticlimax more effective, however, Gulley & Berlo (1956) found no difference between the two. Further consideration to this issue will be given when Thibaut and Walker's (1974) research is reviewed in Section 4.3.2. of this chapter.

In summary, order of presentation in persuasive communication has been investigated from two major perspectives. First, with respect to primacy-recency in which two sources are arguing for a different side on a given issue. Secondly, where the structure of communication of a single source arguing for a single issue is investigated. Primacy effects were reported in early studies but subsequent research showed that both primacy and recency occur depending upon the conditions. Order of a simple communication has been investigated with respect to climax versus anticlimax, explicit versus implicit conclusion. Less attention has been devoted to these issues and the empirical work that there is is not entirely consistent in its findings.

3.3. PSYCHO-LEGAL STUDIES

In the past ten years empirical work setting psychological studies in a legal context has grown dramatically (Tapp, 1976). Much of this work has investigated psychological aspects of judicial processes. Whilst it is beyond the scope and relevance of the present chapter to review the research some indications of the kind of things investigated with respect to jury/juror judgments may be appropriate. With these studies it is important to note that most of the work has used mock juries or jurors and simulated trial proceedings. How much of this detracts from generalizing to real courtrooms is not an issue here -- Bermant *et al*, (1974); Foss (1975) consider this problem in detail and the interested reader is directed to these articles.

Investigations into influences upon juries/jurors have looked at such things as pretrial publicity (Hochberg & Stires, 1973); effect of inadmissable evidence (Wolf & Montgomery, 1977; Sue *et al*, 1973); speech style of defendants and witnesses (Erickson *et al*, 1978; Conley *et al*, 1978); personality characteristics of the defendant (Nemeth & Sosis, 1973). On a more theoretical note models of jury decision-making have been tested by Davis, 1976; Nemeth, 1977; Kerr *et al*, 1976. Perrod & Hastie (1979) provide a review of these models.

The order in which evidence or/and testimony is presented in court can and has been investigated in a number of ways. Perhaps the three most important manipulations concern prosecution/defence

order; witness order in the case for prosecution and defence; finally, the internal order of testimony or evidence given by defence, prosecution or witnesses.

The order of proceedings in criminal trials in Britain and America is of the adversary type. This is structured as follows: prosecution's opening statement. Witnesses for prosecution followed by cross-examination of each in turn by the defence. Next, witnesses for the defence followed by cross-examination of each by prosecution. Finally,the summings up by prosecution, defence and judge.

Research into order effects in a psycho-legal context appear to produce inconsistent findings. Wallace & Wilson (1969) and Wilson (1971) report recency effects, whereas Stone (1969) reports primacy effects. One explanation for these differences might lie in the different experiment procedures. Stone (1969) gets subjects to make three judgments at different points in the "trial". In contrast, the studies reporting recency ask subjects to make judgments only after reading all the material. The effect of forcing mock jurors to make judgments early on in the proceedings may be the reason for obtaining such primacy effects. With the same material and design but only asking subjects to make a final judgment a recency effect might be expected. Stone's procedure might also explain why he failed to find any overall order effects. The above argument is lent credence when one considers the task of jurors. They are encouraged to pay equal attention to arguments appearing in all stages of a trial. This means they are asked, explicitly or implicitly

to suspend judgment until all the evidence has been presented. Stone's (1969) procedure can be seen as dramatically changing this picture. In short, subjects/jurors are not suspending judgment. Given this the study had more value as one on commitment than saying anything substantial about juror decision-making.

In summary, research into prosecution-defence order appears to indicate a tendency towards recency. This contrasts quite markedly with studies of order effect in impression formation, persuasive communciation, etc. There primary effects were most common. Because of the nature of the jurors task, outlined above, this might be expected. Research on impression formation, etc., does not induce or encourage people to wait until all the information has been presented. Indeed some of the explanations of primacy depend upon this assumption. In conclusion, the nature of the trial and the role of the juror explain and may lead us to expect recency effects to be in greater evidence than primary effects.

4.3.2. THIBAUT & WALKER'S RESEARCH PROGRAMME

Thibaut & Walker's (1974) book 'Procedural Justice' is an extended empirical investigation into the nature and impact of procedures in adversary and inquisitorial legal systems. The fundamental question addressed concerned the fairness or how just such procedures are. Using laboratory experiments which involve taking legal roles they report that "one of the clearest findings in our data is that the adversary procedure is judged by all of

our subjects to be the most preferable and the fairest mode of dispute resolution" (p113-115). They claim that their findings apply to conditions of actual litigation, and have no hesitation in making prescriptions for the wholesale adoption of the adversary procedure.

However, the generality of their findings must be treated with greater caution, as Bermant *et al*, (1976) point out. This is because the experiments were conducted with students role-playing and under conditions which do not bear sufficient similarity to actual litigation processes. In what follows their findings will be briefly summarised and a more detailed discussion of the above criticism and others provided.

Thibaut & Walker find the adversary system provides a more extensive search for information and factual material than an inquisitorial system. Here they found that unlawful facts caused decision-makers to give more extreme judgments in the inquisitorial system. With these general findings Thibaut & Walker go on to explore the adversary system in detail. They argue that one of the most important issues is to investigate order effects: "the two-sided character of the adversary system makes the effect of order a fundamental question" (p54). In the inquisitorial system this issue does not arise because the only participant is the decision-maker.

Thibaut & Walker conceptualise potential sources of bias or differences due to order in two ways. Firstly, gross order differences, i.e. whether prosecution or defence go first or last. Secondly, what

they call "internal order", by which they mean the actual presentation of evidence. This relates to a climax or anti-climax order of arguments. This is characterised by Thibaut and Walker as follows: climax is where information and facts are presented in a weak to strong order; anticlimax where material is presented in a strong to weak order.

Chapter 5 of their book "Combating Internal Bias" reports an experimental investigation of this. Walker *et al*, (1972) also provides a report of this research. The study used an imaginary criminal case concerning Adams (the defendant) and Zemp. The case centred around the issue of how justified Adams was in responding violently to an assault by Zemp.

Manipulation of strong and weak material was achieved getting the evidence scaled, using Thurstone's (1959) method of equal-appearing intervals as modified by Young & Cliff (1971). About fifty statements were first divided into lawful or unlawful categories by subjects and then given scale values on 1 - 9 scale. The lawful facts were assigned to the defence, the unlawful to the prosecution.

Three independent variables were manipulated: (i) gross order; (ii) internal order; (iii) role player variation. Only the first two will be considered here as the latter does not have a direct bearing on order effects. Gross order was varied by having role-players read either the lawful followed by the unlawful facts, or unlawful facts followed by lawful ones. Internal order was manipulated in four ways: (a) first attorney presentation sequence of

weak to strong facts, second attorney weak to strong facts (climax-climax); (b) first attorney strong to weak, second attorney strong to weak (anticlimax-anticlimax); (c) climax-anticlimax; (d) anticlimax-climax. Subjects, individually, gave judgments of how lawful or unlawful they thought the defendant's actions were (on a 9-point scale), together with an indication of their degree of certainty about the judgment.

The results showed that the attorney going second obtained more favourable judgments. When prosecution went second judgments were in their favour, when defence went second judgments were in their favour. Thus a recency effect for gross order was obtained. With respect to internal order results showed that for first presentations "climax" was more effective, but only for the prosecution. In the case of the second presentation climax ordering was found to be more effective for both prosecution and defence. Again a recency effect as strong evidence occurring at the end of a presentation i.e. climax, is more influential. As Thibaut and Walker say: "with second presentations, for both prosecution and defence, when strong evidence is presented late in the argument (as in the climax order) it carries greater weight than when it is presented early (as with anticlimax order) -- a recency effect for internal order" (p63). In summary, recency effects are reported for both gross and internal order.

Four serious shortcomings of this study question the degree to

which a realistic simulation of courtroom proceedings is made. These criticisms indicate that one should feel much less confident than Thibaut & Walker in making prescriptions as to how criminal proceedings should be conducted.

The first shortcoming is that the study is unrealistic both with respect to gross and internal order manipulations. To be more realistic any manipulation of order should do more than merely vary the order in which defence and prosecution appear. Such a design falls far short of the actual sequences of events in a courtroom. As described earlier, in any trial there are four distinct phases -- opening statements, case for the prosecution, case for the defence, summings up by prosecution, defence and judge. Given this more complex procedure in the courtroom, compared to that of Thibaut & Walker, we would expect experiments to follow it, varying orders as appropriate, before any kinds of prescriptions could be made with confidence.

The second criticism is that an inadequate simulation is given because Thibaut & Walker do not allow defence/prosecution to come back to one another. There is no cross-examination by either side. Their procedure of assigning unlawful facts to the prosecution and lawful ones to the defence grossly oversimplifies and conflicts[1] with the subtleties of criminal proceedings.

The third criticism concerns the case material itself. Making up and providing a <u>very</u> short summary of testimony, etc., that might be found in a real trial is quite inadequate. Thibaut & Walker present about 500 words of summary. The brevity in itself may be

---

[1] "Unlawful" and "lawful" facts refer to statements indicating guilt or innocence of the defendant <u>not</u> statements permissable or impermissable in a criminal court of law.

one of the reasons for finding order effects. Such a brief summary will have the effect of putting the facts, evidence,etc., into much clear perspective than when the information is divulged over a number of days. These three criticisms, it should be noted, are not specific to Thibaut & Walker, they apply to most of the studies in this area. They are given here because of the strong prescriptive advice made by Thibaut & Walker, and because this study is widely quoted as evidence for order effects in the courtroom.

The fourth, and final, shortcoming is to do with the responses made by subjects. In this study, subjects, it will be recalled, indicated their judgment of lawfulness on a 9-point scale (where 1 was unlawful and 9 lawful). Whilst it is important to discover how order affects judgments in this way it is of even greater importance to find out if such variables affect verdicts (jurors and juries). Thibaut & Walker provide no explicit answer to this issue. They tacitly assume that if judgments on rating scales are affected by gross or internal order than these would translate across to actual verdicts. The grounds for accepting this assumption are not particularly compelling. Significant differences found on rating scale judgments may not mean verdicts are affected. For example, ratings under different orders may both be on the lawful side of the mid-point, here verdicts would also be "lawful" regardless of order. If verdicts are not affected by order there is little reasoning for supposing such a variable to be important in courtroom proceedings. Thibaut & Walker should also have asked subjects to make a verdict, this would have allowed greater generalisability.

In summary, the poor simulation of courtroom proceedings as evidenced by shortcomings in order manipulation, lack of cross-examination, impoverished trial material and inadequate measures means that this study can only be regarded as providing very tentative indications of sources of bias in an actual courtroom. The strong prescriptions Thibaut & Walker make are entirely unwarranted.

The above shortcomings do not mean that the recency effect reported here is uninteresting. The main issue centres around the question of whether the "legal atmosphere", however impoverished itself encourages recency rather than primacy? This relates to the discussion given in the previous section. There, it will be recalled, jurors or subjects in psycho-legal studies were characterised as suspending judgment until all the information has been presented. This discourages primacy effects and recency effects may occur because information appearing later may be more vivid and available from memory when judgment is deliberated upon. A more detailed elaboration of this is provided in Chapter 2 so little more will be said here.

Primacy effects in impression formation, etc, in contrast, are explained as arising because information appearing early on is seen as being stable and later information assimilated to the image formed early on (Anderson & Jacobson, 1965; Anderson & Hubert,1963). However, when subjects know that additional information is going to be presented and are asked to suspend judgment until they have

seen it all the primacy effect disappears (Luchins, 1957), thus validating the characterisation of the task of subjects in psycho-legal experiments given above.

4.4. SUMMARY OF PRIMACY-RECENCY EFFECTS.

From the review of order effects in impression formation, ability attribution, Bayesian judgment, persuasive communication and psycho-legal studies it can be seen no really straightforward guidelines emerge as to when to expect primacy and when to expect recency. Table 4.1 provides a summary of the major studies reviewed here. Inspection of this table reveals that from a purely quantitative perspective more empirical studies have reported primacy rather than recency effects. In impression formation the ordering of lists of six adjectives or more, and short, conflicting paragraphs describing a person produce primacy effects. On the other hand, repeated exposure and short adjective lists nullify order effects. In ability attribution early success/failure influences judgments more than later performance. Large numbers of trials in Bayesian tasks encourage primacy, small numbers encourage recency. Recency in persuasive communication occurs when the material is salient or interesting or moderately unfamiliar. When these conditions are not present or the material is non-salient, or highly interesting or controversial or highly familiar primacy effects are found. In psycho-legal studies recency effects in defence/prosecution order and internal order are reported, primacy only seems to occur when mock jurors are forced to make an early commitment.

Table 4.1. Summary of Main Findings for Order Effects

STUDY	Conditions in which Primacy obtained	Condition in which Recency obtained.	Condition in which No Order Effects obtained.
1. IMPRESSION FORMATION			
ASCH, 1946	List 6 adjectives		
ANDERSON & HUBERT 1963	List 6 adjectives		Repeated exposure
KENDRICK & CONSTANTI 1970	List 6 adjectives		
LUCHINS, 1957	Short paragraphs		
ANDERSON & BARRIOS 1961			Short list adjectives
LUCHINS, 1970			Long time interval
2. ABILITY ATTRIBUTION			
JONES et al, 1967	Early success/ failure		
3. BAYESIAN JUDGMENT			
PETERSON & DUCHARME 1967	Large number trials		
ROBY, 1967	Large number trials		
DALE, 1968	Large number trials		
MILLER, 1968	Large number trials		
PITZ & REINHOLD, 1968		Small number trials	
SHANTEAU, 1970		Small number trials	
4. PERSUASIVE COMMUNICATION			
LUND, 1925	Short arguments		
ROSNOW & GOLDSTEIN 1967	Non-salient material	Salient material	
LANA, 1963a	High interest in topic	Uninteresting topic	
LANA, 1963b	Highly controversial		
LANA, 1961	Highly familiar	Moderately unfamiliar	
5. PSYCHOLEGAL STUDIES			
WALLACE & WILSON 1969		Defense/Prosecution Order	
WILSON, 1971		Defense/Prosecution Order	
STONE, 1969	Early Commitment		
THIBAUT & WALKER 1974		Defense/Prosecution Order Internal Order	

From this it can be seen that research into order effects has progressed from questions of primacy versus recency (as with Lund, 1925) to a detailed inquiry into the conditions under which one or the other effect may be found. However, there is no substantial theoretical model for which primacy or recency may be predicted in different or "real-life" circumstances. As Jones & Goethals(1972) argue "{P}rimacy versus recency is not a gripping critical issue in its own right. Information order, like time, age and number of siblings, is not a conceptually unitary variable" (p28). However, what has emerged from the research, in conjunction with current work on social cognition, is some indication of the underlying cognitive processes which might be involved. While it is beyond the scope of this chapter to provide any detailed review of these it might prove useful to briefly consider them. This will prove useful for the following section.

Jones & Geothals (1972) describe three major processes that are involved in primacy effects and three in recency effects. These are summarised in Table 4.2. Primacy effects are seen to result from attention decrement, discounting and assimilation. Attention decrement is where distraction or fatigue causes people to give undue weight to information appearing early. Discounting is where early information is seen as stable and subsequent conflicting information that may appear later is ignored. Assimilation is where categories are formed early on and later information is assimilated or distorted to fit these categories. These three explanations bear a strong resemblence to the considerations given to theory maintenance and change in Chapter 2, and could be reconceptualised within that framework.

Table 4.2. Summary of Primacy and Recency Explanation (from JONES & GOETHALS, 1972).

ORDER EFFECT	PROCESS	DESCRIPTION
PRIMACY	(i) Attention Decrement (ii) Discounting (iii) Assimilation	Distraction or Fatigue. Early information stable. Categories formed early on.
RECENCY	(i) Recall Readiness (ii) Judgmental Contrast (iii) Content - and context -related hypotheses.	Immediate past better recalled. Short time spans. Progressive changes known.

Recency effects are seen as resulting from the processes of recall readiness (the availability heuristic of Chapter 2?), judgmental contrast and content-and-context-related hypotheses. Recall readiness is simply the fact that immediate past events are better remembered (and more vivid) than ones in the distant past. Judgmental contrast is where later appearing items of information are contrasted with earlier ones when short time spans are involved. Content-and-context-related hypotheses is where people know that progressive changes are likely to occur. This is similar to the suspension of judgment used to explain recency effects in psycho-legal studies discussed earlier.

Although research into order effects is of little theoretical interest in itself they are useful in more substantive areas. Some indication has already been made concerning how the explanations given above may be incorporated into the material discussed in Chapter 2 of this thesis. Furthermore, order effects are regarded as

providing insights into attributional processes (Jones & Goethals 1972). Anderson (1968) uses them to study different models of information processing. By the same token order effects and knowledge of the conditions under which they arise may lead to greater understanding of how outcome knowledge influences people's perceptions and judgments of events. Also, specific information orders may work to reduce or eliminate creeping determinism or hindsight bias. It is to these issues that we now turn.

4.5. ORDER EFFECTS AND HINDSIGHT/FORESIGHT JUDGMENTS.

In what follows the distinction between hindsight bias and creeping determinism, made in Chapter 2, will not be maintained. Hindsight judgments will refer to both, where appropriate. At present there is no reason for supposing information order to affect one without the other, although the results of the experiments reported in PART II of the thesis may cause us to revise this assumption.

In Chapter 3 it was shown that studies which had investigated hindsight judgments could be viewed as falling into three categories. These being experiments which used (a) laboratory material, (b) contemporary socio-political events, (c) factual material. The effect of information order on hindsight judgments can only be asked of the first two categories, in the latter category it is difficult to see how considerations would apply within the format used.

Before looking in more detail at information order and hindsight judgments it is important to show how order may be relevant generally. Here there are two main considerations, firstly, different ordering of information describing some social scenario may act to heighten or reduce hindsight bias or creeping determinism. The exact ways in which this might happen will be considered later. The important point to make here is in the potential application of this to considerations of how historians, policy makers, etc., operate which was discussed in Chapter 1. Specifically, knowing how to reduce bias or the conditions likely to encourage it is prescriptively important. The second reason for looking at information order is a theoretical one. If order is found to have consistent effects on hindsight judgments it may help to resolve some of the theoretical disagreements over the cognitive processes involved. The claim, made at the end of the previous section, that order effects are of little theoretical interest in themselves does not mean they are not of theoretical interest in other areas. As we shall see later in this thesis order of information does provide some important indications to the cognitive processes involved in producing hindsight bias.

4.5.1. INFORMATION ORDER AND HINDSIGHT JUDGMENTS.

Hindsight judgments using laboratory material or contemporary social events may be affected by information order in a number of ways. This section discusses some of these hypotheses. The experiments reported in PART II do not explore all of these hypotheses, only those which appeared, at the time, to be of the greatest significance.

Information bearing on the outcome reported to have happened may be ordered in one of two ways: (i) information which indicates that the outcome is likely to occur may appear first followed by information indicating the occurrence of other outcomes; (ii) information indicating other outcomes may appear first followed by information indicating the reported outcome as likely to occur. To assess how these different orders may affect hindsight judgments we also need to consider the time at which outcome knowledge is given. There are two natural places for providing this, before the event description and after the event description.

It is reasonable to assume that laboratory material does not initiate high interest for subjects and that the social scenario is a moderately unfamiliar one. In fact, the latter is a criterion for selection of social scenarios by Fischhoff (1975). In any case the material is unlikely to be highly controversial or highly familiar to subjects. From the review of order effects, especially in persuasive communication, these considerations might lead us to expect recency effects in foresight judgment.

For hindsight judgments the picture is more complex because of the time at which outcome knowledge is given. If outcome knowledge is given *before the event description* and the information is ordered as in (i) above, then a primacy effect might be expected, resulting in a strong hindsight bias. This may be the case because being told the outcome first and immediately reading information pointing towards that outcome are two conditions under which "assimilation" and "discounting" would be operating. People should, then, regard early

information as stable and hence more diagnostic, with subsequent conflicting information being discounted. Such a presentation should also encourage the early establishment of categories and expectancies into which later information is assimilated or distorted to fit such an outcome. Where outcome knowedge is given before the event description and information order is as in (ii) above less bias should be expected. Here early information will conflict with the outcome people believe has occurred. Hence other outcomes may be seen as more likely than in the previous condition with the result that judgments in hindsight may more closely resemble those in foresight.

Where outcome knowledge is given *after the event descriptions* and information ordered as in (i) above recency effects should occur. This is because of the reasonsgiven earlier concerning the relationship between this material and that of persuasive communications. The recency effect should result, then, in a strong hindsight bias. This is so because if recency effects are operating anyway (foresight judgments in order (i) might be expected to be higher for the outcome believed to have occurred in hindsight than in order (ii)) then giving outcome knowledge at the end of the scenario should serve to increase the likelihood of the outcome pointed to at the end of the event description. In the case where outcome knowledge is given at the end of the scenario and information presented as in order (ii) above, we might expect negligible bias. This is because recency effects may be counteracted by outcome knowledge to produce judgments similar to foresight ones.

Experiments making use of legal material, either real or simulated, demonstrated recency effects. This should apply to hindsight judgments using such material also (providing of course neither side has an overpowering case). This discussion given above should equally apply here.

Experiments using contemporary social issues to investigate hindsight judgments are more difficult to predict order effects for. One problem is that the experiment has no control over when the outcome(s) will occur and what it will be. It may be assumed that such socio-political events would command higher interest, be controversial and be familiar to subjects. Such considerations may lead us to predict primacy effects of varying information order in hindsight. How judgments may be affected in detail is more difficult to hypothesise about and will be left until the particular experiments are described in PART III. This is because the specific contemporary event has to be taken into consideration when making predictions of the effect of information order.

In summary, this section has indicated some ways in which differing orders of information may affect judgments in foresight and, more importantly, in hindsight. Also been discussed is how hindsight bias might be affected (which should equally apply to creeping determinism). This was done in the light of the findings for order effects reviewed earlier in this chapter and the various explanations of primacy-recency effects.

4.5.2. OVERVIEW OF ORDER MANIPULATIONS IN THE EXPERIMENTS IN PART II.

Four of the eight experiments to be described in PART II look at the effects of varying information order upon judgments. Three of these investigate information order with respect to hindsight judgments, one investigates order effects in its own right.

Experiments 1 & 3 (Chapters 5 & 7 respectively), using laboratory material, look at information order and hindsight judgments with respect to unfamiliar social scenarios and 'legal' material. The first experiment manipulated order by an intuitive selection of statements, partially derived from the scaling results of Thibaut and Walker (1974). Experiment 3, because of the findings of order effects in Experiment 1, used a much more rigorous and objective procedure of varying order. Here each statement of an event description was scaled by 20 subjects, scale values determined and information bearing on the outcome presented either first or last in descending strength (similar to climax-anticlimax ordering procedure described earlier). Experiment 5, which used a contemporary social event (The Firemen's Strike of 1977/78), manipulated information order in terms of type of information rather than internal order.

Experiment 4 (Chapter 8), a juror study, investigated order of testimony and witnesses in a long summary of a real courtroom trial. This experiment, which differs from the others because it does not investigate hindsight judgments, was carried out for two main reasons. Firstly, it was noted earlier, in the review of psycho-legal studies

in this chapter, that there were a number of serious shortcomings in this work. The main ones being unrealistic manipulation of order, too short a summary of a case, and little indication of how actual verdicts would be affected. Experiment 4 tries to overcome these problems by manipulating order within the limits of standard courtroom proceedings, providing a more realistic trial summary and by asking mock jurors to give verdicts. Secondly, from looking at order effects within the hindsight/foresight paradigm it is a logical development to look at them on their own with a particularly applied emphasis. Further discussion is given to this aspect later in the thesis when the results of Experiment 4 are discussed.

4.6 SUMMARY

This chapter has reviewed several lines of research dealing with order effects and attempted to summarise the conditions under which primacy or recency may be expected. Also considered was the explanation of the processes involved in producing these effects. The chapter then indicated ways in which order of information may affect hindsight judgments, and the hindsight bias. From this some hypotheses were derived as to which information orders might be expected to encourage the bias and which reduce it. Finally, the chapter briefly described the order manipulations in the experiments reported in PART II of this thesis.

PART II

THE EXPERIMENTS

CHAPTER 5.

EXPERIMENT 1: EFFECTS OF "SURPRISE" INSTRUCTIONS, METHOD OF INFORMATION PRESENTATION AND ORDER ON JUDGMENTS IN HINDSIGHT.

5.1. INTRODUCTION

This experiment examined three issues with respect to hindsight judgments. Firstly, whether "surprise" instructions rather than "ignore", as in Fischhoff (1975a, 1977), would produce hindsight bias. Secondly, the effect of presenting event description information one item at a time rather than continuous prose as in Fischhoff (1975a), and thirdly, the effects of different information orders.

Hindsight judges were given "surprise" instructions instead of "ignore" because of the problem of a demand characteristic interpretation of hindsight bias discussed in Chapter 3. There, it will be recalled, it was argued that asking subjects to "respond as they would have had they not known the outcome" (Fischhoff, 1975a, p293) makes the bias particularly susceptible to such an interpretation. Briefly, the problem centres around giving subjects some piece of information and then asking them to ignore it when making likelihood assessments. They may argue that if the researcher wants them to ignore it why provide it in the first place unless it is what the experiment is about. In view of this, Experiment 1 provided hindsight subjects with "surprise" instructions in which, after being

given outcome knowledge, they were asked to remember that surprising or unexpected events do happen when making likelihood assessments. Such instructions indirectly ask subjects to ignore outcome knowledge because it should lead them to ask themselves how likely the outcome would have been had they not known what happened.

The second issue this experiment investigated, sequential presentation of the event description, was included for two reasons. Firstly, it provides a more realistic simulation of how people go about studying the past than presentation of a continuous prose passage, as in Fischhoff (1975a), does. Typically, people do not have all information to hand at once but sequentially up-date judgments as new information comes to light. Secondly, such a procedure allows a more detailed analysis of how outcome knowledge influences perceptions. If hindsight bias is found it will be possible to determine whether it occurs immediately upon reciept of outcome knowledge or develops as new information comes into view.

The third issue, less central than the previous two, concerns the effects on likelihood assessments of varying the order of information. The experiment uses four scenarios, two taken from Fischhoff (1975a) and two derived from Thibaut & Walker (1975). The latter, which are short summaries of an imaginary criminal case, contain the same information but vary its order. In one (Jury Case A) mostly "unlawful" evidence appears first followed by "lawful" evidence. In the other [1] (Jury Case B) mostly "lawful" followed by "unlawful" evidence is presented. It is hypothesised that hindsight bias will be more evident in Case A than in Case B.

---

[1] For explanation of what is meant by "unlawful" and "lawful" here see footnote on page 121.

This hypothesis is derived from the following considerations: firstly, although it was argued (Chapter 4) that recency effects are dominant where legal material is used, Stone (1969) found a primacy effect when early judgments were made. Secondly, the experiment to be reported here involved early commitment due to the sequential presentation procedure, again likely to produce primacy effects. Given this and that outcome knowledge is provided before the event description a "guilty" outcome with "unlawful" evidence appearing first should encourage hindsight bias. On the other hand the same outcome together with "lawful" evidence appearing first should reduce bias. The detailed reasons for this were discussed in Section 4.5 of Chapter 4 and will not be repeated here.

In summary, the experiment to be reported here attempts to (i) overcome the problem of a demand characteristic interpretation of hindsight bias; (ii) determine whether such bias is present immediately outcome knowledge is given or develops as new information appears; and (iii) investigate how information order influences hindsight judgments.

5.2. METHOD

SUBJECTS

80 subjects, 40 male and 40 female, attending an Open University Summer School took part in the experiment.

PROCEDURE

Subjects were randomly assigned to either a foresight or hindsight

condition and read one of four event descriptions, also assigned at random (with the constraint that equal numbers of men and women were assigned to each experimental treatment). Two of the four scenarios -- "Mrs. Dewar in Therapy" and "The British-Gurkha Struggle" were taken from Fischhoff (1975a), the other two -- "Jury Case A" and "Jury Case B" were constructed from pp128-134 of Thibaut & Walker (1975). In each case subjects were presented with information one item at a time and asked to estimate the likelihood (in percentage form) of each of the possible outcomes specified to a scenario. For Jury Case A and B subjects were asked to assess outcomes concerning the verdicts a hypothetical jury would come to. In all cases subjects were asked to ensure that the four percentages added up to one hundred. After each statement had been revealed it was left open to inspection. Thus the assessment given after statement N had been revealed was made in the context of the previous N-1 items of information that subjects had already seen. After each item of information subjects were asked to assess the likelihood of each of the four possible outcomes. The assessments given to the last statement were taken as reflecting the subjects assessment of the likelihood of the four outcomes in the light of all the information concerning the scenario.

The forty subjects operating in foresight made likelihood assessments without knowing how many items of information they would receive, and without knowing how things turned out in the case given to them. After reading all the information describing the scenario

they were told the outcome and asked, once again, to assess the likelihood of the outcomes presented to them. Before doing this they were reminded that "unexpected or surprising events do happen". This allowed a within-subjects comparison of foresight (F) and hindsight (H1) treatments

Forty subjects were allocated to a second hindsight treatment (H2) in which they were first told the outcome of the scenario randomly allocated to them: reminded that unexpected or surprising events do happen; and then followed the procedure outlined above. This allowed a between-subjects comparison of foresight (F) and hindsight (H2) treatments.

In both conditions (F and H2) subjects spent, on average, between 20 and 25 minutes doing the task. Each subject was run individually with the experimenter revealing the next item of information only after he/she had given likelihood assessments to the few outcomes in the light of the previous statement(s).

INSTRUCTIONS

In the foresight condition the cover sheet of each questionnaire contained the following instructions:

> In this questionnaire we are interested in knowing how people judge the likelihood of possible outcomes of social events. A passage describing an unfamiliar social event will be shown to you. You will be shown statements about the event one at a time. After each statement we would like you to evaluate the probable occurrence of each of the four possible outcomes of the event. When you have

done this the next statement will be shown to you and we would like you to re-evaluate the possible occurrence of each outcome in the light of the new information. Would you please estimate the probabilities in terms of percentages and bear in mind the following information:

(a) when all outcomes seem equally likely, each outcome should be assigned a score of 25%;

(b) after each statement you should write down four scores, which should add up to 100%.

After subjects had read all the information and made likelihood assessments to the four outcomes associated with the final statement they received the following instructions:

> Here is one more item of information I would like you to consider: the actual outcome of this event was....[1]. Please remember that surprising or unexpected events do happen. In the light of this I would now like you to estimate the probability of each of the possible outcomes once more. Thank you.

This is the H1 (Hindsight) treatment.

In the H2 condition the cover sheet of each questionnaire given to subjects was the same as that provided in foresight. Upon reading this subjects were presented with the following information and instructions:

> The first item of information is that the outcome of this event was.....[2] Please remember that unexpected or surprising events do happen. Please estimate the probability of each of the possible outcomes given the information set out below.

---

[1]Depending on the scenario subjects were always given the same outcome. Both the number and actual wording of the outcome were provided. For example, for the scenario "Mrs. Dewar in Therapy" subjects were told that the actual outcome was No.4: continued therapy: improvement.

[2]See footnote[1].

Subjects were then shown the first statement of the scenario and made assessments to the outcomes provided. Subjects had completed their task as soon as the last statement of the scenario had been revealed and assessments made to the outcomes.

STIMULUS MATERIAL

The four event descriptions together with the four possible outcomes specified with each were as follows:

1. MRS. DEWAR IN THERAPY

1. Mrs. Dewar, who was seen once a week, at first was quite resistant to psychotherapy and would evasively say that everything was going along fine and that she had no difficulties.

2. At the fifth session she admitted that she kept having sex dreams about men other than her husband.

3. At the eighth session she confessed, with great self-recrimination and anxiety, that she had had sex relations with one of her uncles when she was sixteen years old.

4. She was apologetic for not bringing out this information sooner.

5. She also said that she felt much better now that she had revealed this.

6. However, her frigidity and menstrual pains continued unabated.

7. At the fifteenth session she confessed that what really concerned her was that, at the age of 13, she had had sexual relations with several boys.

8. Her brother was one of these boys.

9. Each time she now had sex relations with her husband, the memory of what she had done in the past recurred to her.

10. She admitted that she "just didn't want to have anything to do with sex".

11. The therapist opted for a rationalistic approach, trying to convince her that there was no reason why she had to be guilty about her past incestuous relations.

12. The therapist argued that she had engaged in sex relations when she was very young and it is hardly unusual for girls of 13 to be as promiscuous as she was.

OUTCOMES

(1) Terminated therapy; no improvement.

(2) Terminated therapy; improvement.

(3) Continued therapy; no improvement.

(4) Continued therapy; improvement.

2. BRITISH-GURKHA STRUGGLE

1. For some years after the arrival of Hastings as Governor-General of India, the consolidation of British power involved serious war.
2. The first of these wars took place on the Northern frontier of Bengal where the British were faced by the plundering raids of the Gurkhas of Nepal.
3. Attempts had been made to stop the raids by an exchange of lands, but the Gurkhas would not give up their claims to country under British control.
4. Hastings decided to deal with them once and for all.
5. The campaign began in November, 1814. It was not glorious.
6. The Gurkhas were only some 12,000 strong.
7. The Gurkhas were brave fighters, fighting in territory well suited to their raiding tactics.
8. The older British commanders were used to war in the plains where the enemy ran away from a resolute attack.
9. In the mountains of Nepal it was not easy even to find the Gurkhas.
10. The troops and transport animals suffered from the extremes of heat and cold.
11. The British officers learnt caution only after sharp reversals.
12. Major-General Sir D. Octerlony was the one commander to escape from these minor defeats.

OUTCOMES

(1) British victory.

(2) Gurkha victory.

(3) Military stalemate with no peace settlement.

(4) Military stalemate with peace settlement.

3. JURY CASE A.

1. Adams (the defendant) and Zemp had been close friends for years. Recently they began to gamble heavily, and as matters became involved, met at a tavern to discuss their relationship.
2. At the start of the meeting Adams told Zemp they should end their relationship before serious trouble developed between them.
3. After a period of conversation Zemp knocked Adams to the floor and threw an object in his direction.
4. A waiter heard the defendant fall to the floor and scream "I ought to kill you for that".
5. Adams responded by stabbing Zemp in the stomach with a piece of glass.
6. The defendant ran out of the back door after the stabbing.
7. The law provides that it is unlawful to use more force in repelling an attack than a person believes necessary.
8. The defendant (Adams) is 6-feet tall and weights 15 stones; Zemp is 5-feet 8-inches tall and weighs 12 stones.
9. Adams was not injured when he was knocked to the floor.
10. Zemp liked to drink heavily and often became violent when he was drunk.
11. Adams drank only one beer on the night of the stabbing.
12. Before he met Zemp at the tavern the defendant told a friend he hoped that he could settle his dispute with Zemp amiably.
13. One of the waiters told the police he wouldn't mind if Adams never came to tavern again because he sometimes bullied other customers.
14. Zemp told one of the waiters at the tavern he planned to settle his differences with the defendant once and for all.

OUTCOMES OF JURY DELIBERATION

(1) Not guilty.

(2) Guilty of grievous bodily harm -- sentence 5-10 years.

(3) Guilty of using more force than necessary -- sentence 0-2 years.

(4) Hung jury.

4. JURY CASE B.

1. Adams (the defendant) and Zemp had been close friends for years. Recently, they began to gamble heavily, and as matters became involved, met at a tavern to discuss their relationship
2. Before he met Zemp at the tavern the defendant told a friend he hoped that he could settle his dispute with Zemp amiably.
3. Zemp told one of the waiters at the tavern he planned to settle his differences with the defendant once and for all.
4. Zemp liked to drink heavily and often became violent when he was drunk.
5. At the start of the meeting Adams told Zemp they should end their relationship before serious trouble developed between them.
6. After a period of conversation Zemp knocked Adams to the floor and threw an object in his direction.
7. A waiter heard the defendant fall to the floor and scream "I ought to kill you for that".
8. Adams responded by stabbing Zemp in the stomach with a piece of glass.
9. The defendant ran out of the back door after the stabbing.
10. The law provides that it is unlawful to use more force in repelling an attack than a person believes necessary.
11. The defendant (Adams) is 6-feet tall and weighs 15 stones; Zemp is 5-feet 8-inches tall and weighs 12 stones.
12. Adams was not injured when he was knocked to the floor.
13. Adams drank only one beer on the night of the stabbing.
14. One of the waiters told the police he wouldn't mind if Adams never came to the tavern again because he sometimes bullied other customers.

OUTCOMES OF JURY DELIBERATION

(1) Not guilty.

(2) Guilty of grievous bodily harm -- sentence 5-10 years.

(3) Guilty of using more force than necessary -- sentence 0-2 years.

(4) Hung jury.

The two scenarios taken from Fischhoff (1975a) were used because they were found to give the most consistent evidence of hindsight bias. The two scenarios derived from Thibaut & Walker (1975), Jury Case A and Jury Case B, varied information order by first selecting a balanced number of "lawful" and "unlawful" statements from the total number given in their Appendix. Thibaut and Walker obtained this classification for statements by getting judges to scale each according to Thurstone's (1959) method of equal-appearing intervals, hence a scale value was also found for each statement. The statements used, together with their scale values, are reproduced in Appendix 5a. In order to control for length of scenario the number of statements selected was approximately the same as those from Fischhoff (1975a). Order of statements in each of the "Jury Cases" was manipulated by either having the bulk of "lawful" statements appearing first and "unlawful" appearing last or vice versa. This was not done rigorously but in such a way as to preserve the readability of the scenario.

5.3. RESULTS

To discover how outcome knowledge, sequential presentation and information order (in the two "Jury Cases") influenced likelihood assessments, comparisons were made using non-parametric tests.[3] The Wilcoxon Matched Pairs Signed Ranks Test was used for repeated measures comparisons, the Mann-Whitney U-test for independent measures (Siegel, 1956). One-tailed tests were used in foresight/hindsight

---

[3] Of the seven experiments investigating foresight/hindsight judgments reported in this thesis all except Experiment 8 used non-parametric test. The reasons for using such tests is discussed in Appendix 12.

and information order comparisons, two-tailed tests in information presentation comparisons.

(a) Hindsight/Foresight

To obtain an overall picture of how judgments made in hindsight differed from those made in foresight the analysis first pooled results within conditions across the fours scenarios. This was done for both outcomes reported to have occurred and not reported to have occurred. In both cases this was achieved by using a method similar to Fischhoff & Beyth (1975) and modified in light of the discussion given in Chapter 3.

For outcomes reported to have occurred the analysis proceded as follows: the median assessments for each of the four scenarios in the foresight condition after all statements had been revealed was determined. For each scenario the median was compared with the response of each subject in the H2 condition. In order to decide whether a subject was hypothesis supporting (+) or non-hypothesis supporting (-) the following coding scheme was used:-

(+) assigned if the H2 raw score was greater than the foresight median for the particular scenario;

(-) assigned if the H2 raw score was less than or equal to the foresight median for the scenario.[4]

Table 5.1 summarises the results of this analysis. The table shows

---

[4] Fischhoff & Beyth (1975a) were shown in Chapter 3 to use a procedure which omitted scores when foresight equalled hindsight.. The discussion there considered this invalid, hence "draws" are considered non-hypothesis supporting.

Table 5.1 Number of hypothesis and non-hypothesis supporting subjects in each scenario for the outcome reported to have occurred.

SCENARIO AND OUTCOME	(+)	(-)
Mrs. Dewar in therapy: continued therapy -- improvement (No.4)	3	7
British-Gurkha Struggle -- British Victory (No.1)	7	3
Jury Case A -- Guilty of using more force than necessary (No.3)	7	3
Jury Case B -- Guilty of using more force than necessary (No.3)	8	2
Total	25	15
%	62.5	37.5

that across the four scenarios 62.5% of subjects were hypothesis supporting and 37.5% non-hypothesis supporting. This is not a significant difference however (Binomial test $Z = 1.42$, $p < 0.078$, one-tailed). Fischhoff & Beyth (1975), using their procedure, find over 69% to be hypothesis-supporting (this drops to around 61% if the above procedure is used -- see Table 3.1 for further details).

A similar procedure for outcomes not reported to have occurred was used. The only difference was that the following coding scheme was used in place of the one described earlier:-

(+) assigned if H2 raw score was less than the foresight median for that particular scenario.

(-) assigned if H2 raw score was greater than the foresight median for that particular scenario.

Table 5.2 summarises the results of this analysis, and shows that

Table 5.2 Number of hypothesis-supporting and non-supporting subjects in each scenario for outcomes reported not to have occurred.

SCENARIO AND OUTCOME		(+)	(-)
Mrs. Dewar in therapy	No.1	6	4
	No.2	5	5
	No.3	2	8
	Total	13	17
British-Gurkha	No.2	6	4
	No.3	8	2
	No.4	8	2
	Total	22	8
Jury Case A	No.1	-	10
	No.2	7	3
	No.4	-	10
	Total	7	23
Jury Case B	No.1	-	10
	No.2	10	-
	No.4	-	10
	Total	10	20
	Grand Total	52	68
	%	43.3	56.6

across the four scenarios only 43.3% of estimates given in hindsight were hypothesis-supporting, whilst 56.6% were non-hypothesis supporting. Fischhoff & Beyth (1975) report over 55% to be hypothesis-supporting (this drops to around 50% if "draws" are included -- see Table 3.1 for more on this).

To discover how outcome knowledge together with the "surprise" instructions influenced judgments in each of the four scenarios further foresight/hindsight comparisons were made. This was done using the Mann-Whitney U-test in F-H2 comparisons, and Wilcoxon

Table 5.3 Median probability estimates given to the reported outcome after all information presented.

CASE	OUTCOME	CONDITION		
		FORESIGHT	HINDSIGHT(2)	HINDSIGHT(1)
1. Mrs. Dewar in therapy	Continued therapy -- improvement	70.0	55.0 p<0.02	82.5 p<0.01
2. British-Gurkha Struggle	British victory.	17.5	55.0 p<0.05	57.5 p<0.05
3. Jury Case A	Guilty of using more force than necessary -- sentenced 2yrs.	32.5	52.0 NS	75.0 p<0.05
4. Jury Case B	Guilty of using more force than necessary -- sentenced 2 yrs.	3.5	55.0 p<0.025	30 p<0.025

in the F-H1 comparisons.

Table 5.3 summarises the median likelihood assessments given

to the reported outcomes after all items of information had been presented in F, H1 and H2 treatments replicated. Fischhoff's (1975a) results with respect to the "British-Gurkha Struggle" ($p<0.05$, Mann-Whitney $U=26$) and showed that a similar bias was evident in "Jury Case B" ($p<0.025$, Mann-Whitney $U=23$). "Jury Case A" produced no statistically reliable difference between F and H2 treatments but the scenario "Mrs. Dewar in Therapy" produced data opposite to that in Fischhoff's research ($p<0.02$, Mann-Whitney $U=18.5$, two-tailed).

Table 5.3 also shows that data from F and H1 treatments (repeated measures) led to higher assessments for reported outcomes in hindsight than foresight for each of the four scenarios ("Mrs. Dewar in Therapy", $p<0.01$, Wilcoxon $T=1$, $N=8$; "British-Gurkha Struggle", $p<0.005$, Wilcoxon $T=0$, $N=8$; "Jury Case A", $p<0.05$, Wilcoxon $T=0$, $N=5$; "Jury Case B", $p<0.025$, Wilcoxon $T=0$, $N=6$).

A similar analysis is not given here for outcomes reported not to have occurred because little difference between foresight and hindsight (H2) was found in the pooled results, as indicated in Table 5.2. Appendix 5B summarises, for each scenario and each outcome, all the median percentages derived after each item of information was presented. From this it can be seen that many hindsight assessments are similar to foresight ones for outcomes reported not to have occurred.

(b) Hindsight Bias and Sequential Presentation of Information.

The sequential method of information presentation allows us to determine whether hindsight bias, occurs immediately outcome knowledge is given or develops as new information comes into view. To investigate this two sets of analyses were conducted. First, between foresight and hindsight (H2) assessments given after the first statement has been presented.

Table 5.4 Median probability estimates given to the reported outcome after first item of information.

CASE	OUTCOME			SIGNIFICANCE
		FORESIGHT	HINDSIGHT(2)	
1. Mrs. Dewar in therapy	Continued therapy -- improvement	60.0	22.5	$p<0.05$
2. 'British-Gurkha struggle	British victory	45.0	30.0	NS
3. Jury Case A	Guilty of using more force than necessary -- sentenced 2 yrs.	25.0	25.0	NS
4. Jury Case B	Guilty of using more force than necessary -- sentenced 2 yrs.	13.0	25.0	NS

Second, between assessments given to the first and last statement of information in F and H2 for each scenario.

Table 5.4 summarises the median assessments given to the reported outcome after presentation of the first statement in treatments F and H2. From this it can be seen that for "Jury Case B" the effect of outcome knowledge was to raise the perceived probability, but not significantly so ($p>0.05$, Mann Whitney $U=43$). For "Jury Case A" there was no difference between F and H2. In the other two scenarios the effect of outcome knowledge was to lower the perceived probability of the reported outcome. The difference was only significant for "Mrs. Dewar in Therapy" ($p<0.05$, Mann Whitney $U=22$, two-tailed), not in the case of the "British-Gurkha Struggle" ($p>0.05$, Mann Whitney $U=43.5$). From this it appears that differences between F and H2 do not appear immediately outcome knowledge is given.

Table 5.5 Median probability estimates given to the reported outcome after first and last items of information.

CASE	CONDITION	STATEMENT		SIGNIFICANCE
		FIRST	LAST	
1. Mrs. Dewar in Therapy	Foresight(F)	60.0	70.0	NS
	Hindsight(H2)	22.5	50.0	$p < 0.05$
2. British-Gurkha Struggle	Foresight(F)	45.0	17.5	$p < 0.05$
	Hindsight(H2)	30.0	55.0	NS
3. Jury Case A	Foresight(F)	25.0	32.5	$p < 0.05$
	Hindsight(H2)	25.0	52.0	NS
4. Jury Case B	Foresight(F)	13.0	3.5	NS
	Hindsight(H2)	25.0	55.0	$p < 0.01$

Table 5.5 summarises the median assessments given to the reported outcomes after presentation of the first and last items of information in F and H2 respectively. From this it can be seen that in all four scenarios judges operating in hindsight perceived the reported outcome as more likely at the end than the beginning. Only two of these were significantly different though ("Mrs. Dewar in Therapy", p 0.05, Wilcoxon T=9.5, N=10; Jury Case B, p 0.01, Wilcoxon T=0, N=8). In contrast, judges operating in foresight

Table 5.6 Median Probability Estimates of the Four Outcomes in Jury Cases A & B given in Foresight and Hindsight(H2).

OUTCOME	CONDITION	SCENARIO		SIGNIFICANCE
		JURY CASE A	JURY CASE B	
1. Not Guilty	Foresight(F)	0	0	NS
	Hindsight(H2)	0	0	NS
2. Guilty of causing grievous bodily harm	Foresight(F)	37.5	87.5	$p < 0.05$
	Hindsight(H2)	17.5	30.0	NS
3. Guilty of using more force than necessary	Foresight(F)	32.5	3.5	$p < 0.025$
	Hindsight(H2)	52.0	55.0	NS
4. Hung Jury	Foresight(F)	0	0	NS
	Hindsight(H2)	0	7.5	NS

revised their estimates downwards in some cases ("The British-Gurkha Struggle", $p<0.05$, Wilcoxon T=6, N=9; "Jury Case B", $p>0.05$, Wilcoxon T=10, N=6) and upwards in others ("Mrs. Dewar in Therapy", $p>0.05$, Wilcoxon T=9, N=9; "Jury Case A", $p<0.05$, Wilcoxon T=4, N=9). In foresight, then, there is no consistent pattern, in hindsight there is.

(c) Information Order in Jury Case A and B.

To discover how the two different information orders in Jury Case A and B affected judgments two sets of analyses were performed on estimates given after the last statement. Firstly, between judgments in Jury Case A and B for each outcome. Table 5.6 summarises the medians and results of this analysis. Figures 5.1 and 5.2 graphically depict the medians for outcomes 2 and 3 respectively. From the table it can be seen that for outcome 2 (guilty of causing grievous bodily harm) judges in Jury Case B found it to be much more likely than those in Jury Case A ($p<0.05$, Mann Whitney U=27). For outcome 3 (guilty of using more force than necessary) the opposite was found, i.e. subjects in "A" believed it much more likely than those in "B" ($p<0.025$, Mann-Whitney U=21), between hindsight assessments in the two cases produced no significant differences. The hypothesis that hindsight bias would be more evident in the Case A than Case B because of primacy effects, was not confirmed: the pattern of results indicate recency effects to be operating.

The second analysis was between foresight and hindsight assessments in each case for each outcome. For the outcome reported to have occurred (guilty of using more force than necessary) previous analysis (see Table 5.3) has already shown hindsight assessments to be significantly greater only for Jury Case B. For outcomes not reported to have occurred only outcome 2 in Jury Case B was found to be significantly greater in foresight than hindsight ($p<0.03$, Mann Whitney U=23.5).

Figure 5.1 Median likelihood assessments in foresight and hindsight (H2) for the outcome "Guilty of grievous bodily harm.

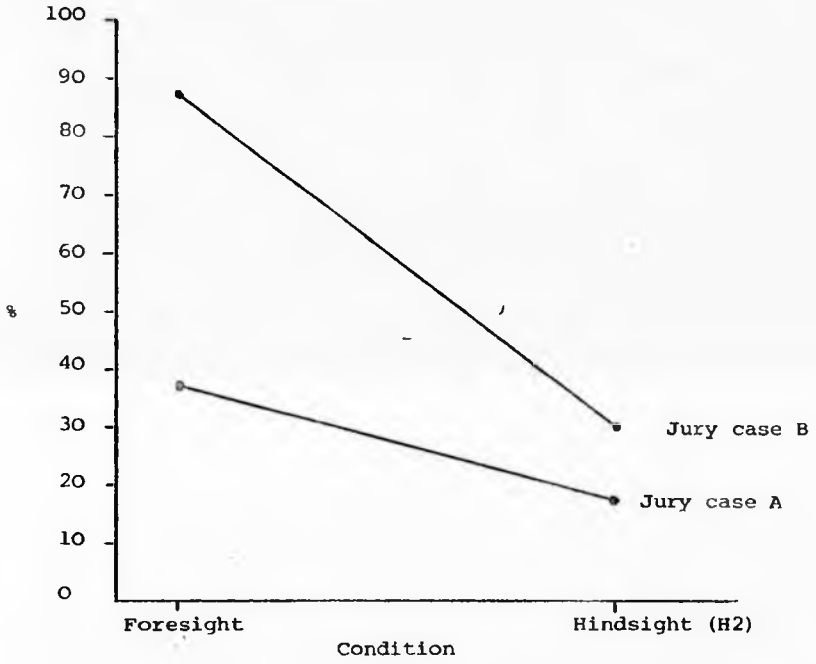
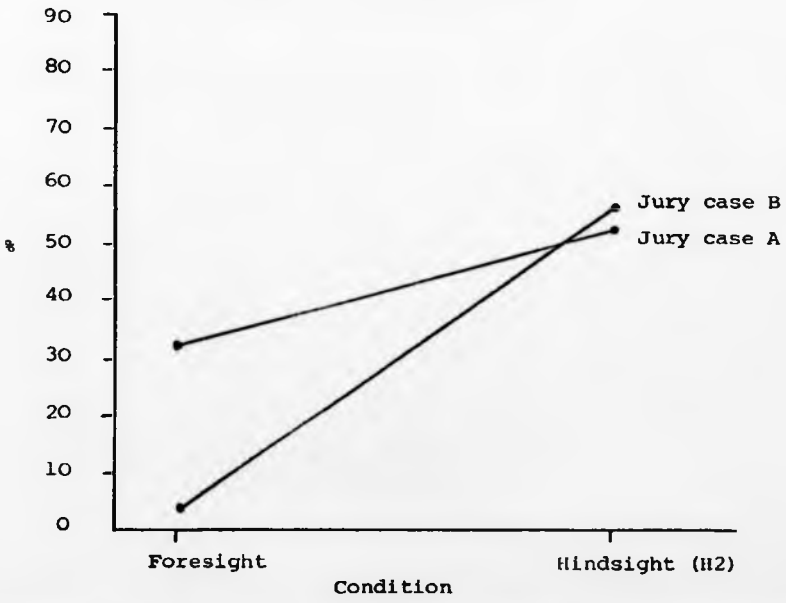


Figure 5.2 Median likelihood assessments in foresight and hindsight (H2) for the outcome "Guilty of using more force than necessary" (this is the outcome reported to have occurred).



5.4. DISCUSSION

The first issue this experiment was designed to explore was whether outcome knowledge would lead subjects to exhibit hindsight bias when they were provided with "surprise" rather than "ignore" instructions. Strong support for hindsight bias was not found here. Reporting an outcome of a scenario produced a consistent increase in its likelihood whenever within subjects comparisons were made, thus replicating and extending the findings of Fischhoff & Beyth (1975a). However, between subjects comparisons produced much less consistent results. This was shown in two ways: firstly, an overall picture of how foresight and hindsight judgments differed, made by pooling across scenarios, found less support than that reported by Fischhoff & Beyth, (1975a). The experiment reported in this Chapter found only 62.5% of subjects to be hypothesis-supporting for outcomes reported to have occurred, compared with 69% by Fischhoff & Beyth (1975a). Furthermore, only 43.3% of subjects were hypothesis-supporting for outcomes reported not to have occurred, Fischhoff and Beyth's figure being over 55%. Secondly, turning to a consideration of F-H2 comparisons for each scenario, those taken from Fischhoff (1975a) replicated previous findings with respect to "The British-Gurkha Struggle", but reversed previous findings for "Mrs. Dewar in Therapy". Considering the two scenarios derived from Thibaut and Walker (1975, pp128-134), F1H2 comparisons were in the direction predicted by Fischhoff's research, but only one (Jury Case B) was statistically significant. Taken together, these findings suggest hindsight bias not to be a particularly strong or reliable phenomenon.

The second issue concerned just how outcome knowledge influences perceptions and how hindsight bias occurs. The sequential presentation procedure allowed this to be investigated. For three scenarios it was found that F and H2 assessments were not significantly different after the first statement had been presented. For the scenario, "Mrs. Dewar in Therapy", however, H2 assessments were significantly less than foresight ones. Together these findings demonstrate that if hindsight bias is going to occur it will not do so immediately upon receipt of outcome knowledge. Comparisons between assessments assigned to the first and last statements of event descriptions in foresight and hindsight (H2) indicated a consistent pattern of results in the latter treatment. For each scenario it was found that all assessments given to the last statement were higher than those given to the first; indicating that if hindsight bias is going to occur it does so only as new information comes to light. However, because only weak support for hindsight bias was found as indicated earlier, it must be concluded that the above influence of outcome knowledge whilst consistent, may not be great enough to lead to reliable bias.

The third issue was how foresight/hindsight judgments would be affected by different orders of information in "Jury Case A" and "Jury Case B". It was hypothesised that hindsight judges in Case A would exhibit greater bias than such judges in Case B. The opposite was found, indicating recency, not primacy, effects to be operating. This can be seen from inspection of Figures 5.1 and 5.2. Both graphs show outcome knowledge to have greater influence over judgments in

Jury Case B rather than Jury Case A. In Figure 5.1, which plots median assessments for Outcome 2, hindsight assessments are significantly less for Case B but not for Case A. This is in the expected direction if hindsight bias is operating (i.e. assessments less in hindsight for outcomes reported not to have occurred). Figure 5.2 which plots medians for the outcome reported to have occurred, demonstrates an equivalent pattern. Hindsight assessments are significantly greater than foresight ones in Jury Case B but not in Jury Case A. The greater influence of outcome knowledge in "B" can only be attributable to the different information orders. In A the order was mostly "unlawful" followed by "lawful" statements, and the reverse of this in B. The above results indicate that information appearing last had greater influence over hindsight assessments. Explaining why bias was found in "B" for Outcomes 2 and 3 and not in "A" for the same outcomes. The implications of these results are that hindsight bias may be reduced if information bearing on other outcomes appears last.

Why should recency effects of the sort described above be in evidence rather than primacy ones as expected from the discussion given in the introduction to this experiment? No indication of this can be found in the results. Consideration of the material provides the only indication. In Chapter 4 it was shown, on the whole, that recency effects dominated where "legal" material was used. A main reason for this was a "suspension of judgment" hypothesis. Results

from foresight comparisons indicate such a principle to be operating here. Also, in Chapter 4, it was indicated that early commitment encouraged primacy effects, this was demonstrated by Stone (1969). Commitment in the experiment reported here differed from Stone's in that subjects made verdicts in his study and likelihood assessments in this. This difference may account for the recency effects found here as commitment is less strong.

Overall the findings discussed here lend only partial support for Fischhoff's claim that events seen in hindsight are perceived as more inevitable than when those same events are seen in foresight. Apart from the above reasons for this three other considerations need to be discussed.

First, the "ignore" instructions of Fischhoff may indeed suffer from demand characteristic problems, the use of "surprise" instructions is less susceptible to this and so may provide a more accurate picture of how judgments made in hindsight differ from those in foresight. However, it has to be acknowledged that such instructions may lead people to be too conservative when making judgments.

Second, due to the problem of the "law of small numbers" (Tversky and Kahneman, 1971) discussed in Chapter 2, it is possible that less reliable or consistent results found in this experiment were due to sample size being approximately half that of Fischhoff, (1975a); Fischhoff & Beyth, (1975a). This is dealt with in some detail in Appendix 13 as it is an issue relevant to other experiments reported in

from foresight comparisons indicate such a principle to be operating here. Also, in Chapter 4, it was indicated that early commitment encouraged primacy effects, this was demonstrated by Stone (1969). Commitment in the experiment reported here differed from Stone's in that subjects made verdicts in his study and likelihood assessments in this. This difference may account for the recency effects found here as commitment is less strong.

Overall the findings discussed here lend only partial support for Fischhoff's claim that events seen in hindsight are perceived as more inevitable than when those same events are seen in foresight. Apart from the above reasons for this three other considerations need to be discussed.

First, the "ignore" instructions of Fischhoff may indeed suffer from demand characteristic problems, the use of "surprise" instructions is less susceptible to this and so may provide a more accurate picture of how judgments made in hindsight differ from those in foresight. However, it has to be acknowledged that such instructions may lead people to be too conservative when making judgments.

Second, due to the problem of the "law of small numbers" (Tversky and Kahneman, 1971) discussed in Chapter 2, it is possible that less reliable or consistent results found in this experiment were due to sample size being approximately half that of Fischhoff, (1975a); Fischhoff & Beyth, (1975a). This is dealt with in some detail in Appendix 13 as it is an issue relevant to other experiments reported in

from foresight comparisons indicate such a principle to be operating here. Also, in Chapter 4, it was indicated that early commitment encouraged primacy effects, this was demonstrated by Stone (1969). Commitment in the experiment reported here differed from Stone's in that subjects made verdicts in his study and likelihood assessments in this. This difference may account for the recency effects found here as commitment is less strong.

Overall the findings discussed here lend only partial support for Fischhoff's claim that events seen in hindsight are perceived as more inevitable than when those same events are seen in foresight. Apart from the above reasons for this three other considerations need to be discussed.

First, the "ignore" instructions of Fischhoff may indeed suffer from demand characteristic problems, the use of "surprise" instructions is less susceptible to this and so may provide a more accurate picture of how judgments made in hindsight differ from those in foresight. However, it has to be acknowledged that such instructions may lead people to be too conservative when making judgments.

Second, due to the problem of the "law of small numbers" (Tversky and Kahneman, 1971) discussed in Chapter 2, it is possible that less reliable or consistent results found in this experiment were due to sample size being approximately half that of Fischhoff, (1975a); Fischhoff & Beyth, (1975a). This is dealt with in some detail in Appendix 13 as it is an issue relevant to other experiments reported in

PART II of this thesis, rather than one specific to the experiment reported in this chapter.

Third, it may be the case that hindsight bias is mediated by the position of outcome knowledge relative to other items of information in the scenario. H1 subjects received outcome knowledge at the end of the event description, as in Fischhoff (1975a), and saw the reported outcomes as more probable than in foresight. H2 subjects were told what had happened before they received other information and produced the expected pattern of results in only two of the four cases.

A second experiment was designed, therefore, to investigate the effects of the position of outcome knowledge when subjects read continuous prose passages as in Fischhoff (1975a).

CHAPTER 6.

EXPERIMENT 2: POSITION OF OUTCOME KNOWLEDGE AND HINDSIGHT JUDGEMENTS

6.1 INTRODUCTION

The results of the previous experiment indicated that hindsight bias may not be a particularly strong or consistent phenomenon. One reason for this was that two variables were found to affect it, firstly, it may be mediated by the position of outcome knowledge in relation to the event description. Secondly, certain information orders may act to reduce or encourage it. In contrast to the previous experiment the one to be reported here investigated creeping determinism rathern than hindsight bias.

There are two reasons for this. Firstly, the lack of strong support for hindsight bias makes it of interest to attempt tc replicate Fischhoff's (1975a) results with respect to creeping determinism. Secondly, as Fischhoff (1975a) reports, creeping determinism to be a stronger phenomenon than hindsight bias we may be able to determine more precisely the effects of position of outcome knowledge and information order.

It will be recalled from Chapter 3 that a distinction was made between creeping determinism and hindsight bias. The latter was said to occur only when subjects were given instructions which, either explicitly or implicitly, asked them to ignore outcome knowledge

CHAPTER 6.

EXPERIMENT 2: POSITION OF OUTCOME KNOWLEDGE AND HINDSIGHT JUDGEMENTS

6.1 INTRODUCTION

The results of the previous experiment indicated that hindsight bias may not be a particularly strong or consistent phenomenon. One reason for this was that two variables were found to affect it, firstly, it may be mediated by the position of outcome knowledge in relation to the event description. Secondly, certain information orders may act to reduce or encourage it. In contrast to the previous experiment the one to be reported here investigated creeping determinism rathern than hindsight bias.

There are two reasons for this. Firstly, the lack of strong support for hindsight bias makes it of interest to attempt tc replicate Fischhoff's (1975a) results with respect to creeping determinism. Secondly, as Fischhoff (1975a) reports, creeping determinism to be a stronger phenomenon than hindsight bias we may be able to determine more precisely the effects of position of outcome knowledge and information order.

It will be recalled from Chapter 3 that a distinction was made between creeping determinism and hindsight bias. The latter was said to occur only when subjects were given instructions which, either explicitly or implicitly, asked them to ignore outcome knowledge

(i.e. attempt to reconstruct a foresight state) when making judgements. Creeping determinism, on the other hand, was said to occur when subjects, provided with outcome knowledge and an event description, were asked to make judgements in the light of all the information they had been given. Fischhoff (1975a) finds creeping determinism to be a stronger phenomenon than hindsight bias. For example, pooling his results from Tables 1 and 3 (pp. 293 and 295 respectively) for the outcome reported to have occurred we find that the median foresight/hindsight difference is 14.2% and 9.5% for creeping determinism and hindsight bias respectively.

The experiment to be reported here investigated three hypotheses: that creeping determinism would (1) be more in evidence when outcome knowledge was given after the event description than before; (2) that confidence in the judgements made by subjects would be greater in hindsight than foresight; (3) that information appearing last which indicates that the reported outcome will occur will have a stronger influence on judgements than information which appeared first.

The first hypothesis is an extension of the indications in the previous experiment that outcome knowledge given at the end of an event description may induce greater hindsight bias than if it were given at the beginning. The hypothesis is lent further credence by the recency effects reported in the previous experiment. If subjects are instructed to make judgements in the light of all information, as in those for creeping determinism, the existence of recency effects would serve to make the reported outcome appear more inevitable

when outcome knowledge is given last rather than first.

The second hypothesis also derives from Fischhoff's general finding that outcome knowledge leads people to perceive that outcome as more inevitable than if they did not know the outcome. This is so because it follows from this that we would expect people to feel more confident in their judgements in hindsight. Indirect support for this hypothesis can be found in a number of historical writings. For example, Wohlsletter (1962) comments:

> " . . . it is much easier after the event to sort out the relevant from the irrelevant signals. After the event, of course, a signal is always crystal clear; we can now see what disaster it was signalling, since the disaster has occurred." (p. 387)

The third hypothesis is made as a result of the findings in the previous experiment, where, as discussed above, recency effects were found. There is no obvious reason why this should not apply when creeping determinism rather than hindsight bias is being investigated.

The experiment to be reported here uses the same scenarios and outcomes as in the previous one. It differs in that subjects were presented with continuous prose passages rather than the sequential presentation of information method. This was done because of the focus upon replication and extension of Fischhoff (1975a). Consequently it was thought more appropriate to follow his design in this respect.

6.2. METHOD

SUBJECTS

120 undergraduate students at the University of Warwick acted as subjects in this experiment.

PROCEDURE

40 subjects were randomly assigned to a Foresight (F) condition: 40 to a hindsight condition (HB) in which they were told what happened before reading the event description (this is equivalent to the H2 condition in Experiment 1); and 40 to a hindsight condition (HA) in which they were told the outcome after reading the details of the scenario. In each condition subjects read one (and only one) of the four scenarios used in Experiment 1. Subjects in all conditions were asked to make two sets of judgements, first to assess the likelihood (in percentage form) of each of the possible outcomes specified to a scenario; secondly, to indicate how confident they felt about their assessments. They did this on a scale where 0 was "unsure" and 5 was "very confident". Likelihood assessments, as in Experiment 1, given to the four possible outcomes of a scenario had to add up to 100%.

INSTRUCTIONS

The instructions given to both foresight and hindsight subjects that appeared on the coversheet of the questionnaire ran as follows:

> In this questionnaire we are interested in knowing how people judge the likelihood of possible outcomes of social events. A passage describing an unfamiliar social event will be shown to you. We would like you to evaluate the probability of occurrence of each of the four possible outcomes of the event in the light of the information appearing in the passage.
>
> Would you please estimate the probabilities in terms of percentages and ensure that the four scores add up to 100%
>
> Please look at the questionnaire sheet (p.3), before you read

> the event description, so that you are acquainted with the outcomes.

These instructions are similar to those of Fischhoff (1975) in Experiment 1 (p. 289), which are reproduced in Appendix 6A. After reading the event description as appropriate to the condition, subjects were given the following instructions to read before making judgements:

> In the light of the information appearing in the passage what was the probability of occurrence of each of the four possible outcomes listed below. (The probabilities should sum to 100%).

After this subjects were asked to indicate their degree of confidence in their judgements on a scale labelled as indicated above.

STIMULUS MATERIAL

The four scenarios used in this experiment ("Mrs. Dewar in Therapy"; "The British-Gurkha Struggle"; "Jury Case A"; "Jury Case B") were the same as those used in Experiment 1. The only difference was that numbers to each sentence had been omitted and the sentences combined to form a single continuous prose passage. In the conditions where outcome knowledge was given it appeared separately either at the beginning or end of the paragraph. For example, in HA (hindsight, outcome knowledge after the event description) the "British-Gurkha Struggle" was presented to subjects as follows:

2. BRITISH GURKHA STRUGGLE

> For some years after the arrival of Hastings as Governor-General of India, the consolidation of British power

involved serious war. The first of these wars took place on the Northern Frontier of Bengal where the British were faced by the plundering raids of the Gurkhas of Nepal. Attempts had been made to stop the raids by an exchange of lands, but the Gurkas would not give up their claims to country under British control, and Hastings decided to deal with them once and for all. The campaign began in November, 1814. It was not glorious. The Gurkhas were only some 12,000 strong; but they were brave fighters, fighting in territory well-suited to their raiding tactics. The older British commanders were used to war in the plains where the enemy ran away from a resolute attack. In the mountains of Nepal it was not easy even to find the Gurkhas. The troops and transport animals suffered from the extremes of heat and cold, and the officers learnt caution only after sharp reversals. Major-General Sir D. Octerlony was the one commander to escape from this minor defeat.

The actual outcome of this event was a British Victory. (Outcome No. 1 on questionnaire sheet).

6.3 RESULTS

(1) Position of Outcome Knowledge and Creeping Determinism

To discover the effect of (a) outcome knowledge and (b) position of that information relative to the event description, upon judgements two sets of analyses were performed. First, to achieve an overall picture the same pooling across studies procedure as described in Experiment 1 was used. Second, individual comparisons between foresight and hindsight judgements for each scenario was made using the Mann Whitney U test.

Tables 6.1 and 6.2 summarise the results of the "pooled analysis" for outcomes reported and not reported to have occurred respectively. Table 6.1 shows that when outcome knowledge was given before the event description only 37.5% of subjects exhibited creeping determinism (i.e. were hypothesis supporting), the majority, 62.5%, being non-hypothesis

supporting. The Table also shows that when outcome knowledge is provided after the event description 42.5% of subjects exhibited creeping determinism. These figures for creeping determinism are much lower than those reported by Fischhoff.

Table 6.2 shows that for outcomes reported not to have happened 40% were hypothesis supporting when outcome knowledge was given before the event description and 49.2% of assessments when outcome knowledge was given at the end. These are higher than average percentages obtained for outcomes reported to have occurred, which is opposite to Fischhoff's findings. Looking at the individual scenarios, Table 6.1 shows that outcome knowledge given after the event description is more likely to result in creeping determinism than if it is given before in the scenarios "The British-Gurkha Struggle" and "Jury Case A".

Turning now to individual comparisons, Table 6.3 summarises the median assessments made in the foresight and two hindsight conditions for each outcome in each scenario. The Table shows that, for the outcomes reported to have occurred, hindsight after (HA) judgements were in the predicted direction for the "British-Gurkha Struggle" ($P < 0.025$, Mann Whitney $U = 19.5$) and "Jury Case B" ($P < 0.1$, Mann Whitney $U = 29$). Hindsight (after) assessments were in the opposite direction to that expected by the creeping determinism hypothesis, neither being significantly so however. For outcomes reported to have occurred in the hindsight before (HB) condition no assessments were significantly different from foresight ones. For outcomes which did not occur foresight/hindsight comparisons yielded two significant

Table 6.1 Number of hypothesis supporting and non-supporting subjects in each scenario for outcomes reported to have occurred

SCENARIO AND OUTCOME	CONDITION			
	HINDSIGHT BEFORE		HINDSIGHT AFTER	
	(+)	(-)	(+)	(-)
Mrs. Dewar in Therapy: continued Therapy - improvement (No.4)	2	8	3	7
British-Gurkha Struggle: British Victory (No.1)	6	4	8	2
Jury Case A: Guilty of using more force than necessary (No.3)	2	8	1	9
Jury Case B: Guilty of using more force than necessary (No.3)	5	5	5	5
Total	15	25	17	23
%	37.5	62.5	42.5	57.5

Table 6.2 Number of hypothesis-supporting and non-supporting assessments in each scenario for outcomers reported not to have occurred

SCENARIO	OUTCOME NUMBER	CONDITION			
		HINDSIGHT BEFORE		HINDSIGHT AFTER	
		(+)	(-)	(+)	(-)
Mrs. Dewar in Therapy	1	6	4	5	5
	2	3	7	3	7
	3	2	8	5	5
	TOTAL	11	19	13	17
The British-Gurkha Struggle	2	6	4	6	4
	3	4	6	9	1
	4	4	6	4	6
	TOTAL	14	14	19	11
Jury Case A	1	4	6	3	7
	2	4	6	2	8
	4	0	10	3	7
	TOTAL	8	22	8	22
Jury Case B	1	4	6	6	4
	2	8	2	7	3
	4	3	7	6	4
	TOTAL	15	15	19	11
GRAND TOTAL		48	72	59	61
%		40.0	60.0	49.2	50.8

Table 6.3 Summary of Median Assessments for each Scenario and each Outcome given in Foresight and the two hindsight conditions

SCENARIO AND OUTCOME	CONDITION		
	FORESIGHT	HINDSIGHT BEFORE	HINDSIGHT AFTER
Mrs. Dewar in Therapy			
1. Terminated Therapy - no improvement	20	10	17.5
2. Terminate Therapy - improvement	10	10	10
3. Continued Therapy - no improvement	17.5	20	22.5
4. Continued Therapy - improvement	50	50	37.5
The British-Gurkha Struggle			
1. British Victory	12.5	17.5	37.5**
2. Gurkha Victory	50	35	37.5+
3. Military Stalemate with peace settlement	22.5	27.5	5.0
4. Military Stalement with no peace settlement	10	10	10
Jury Case A			
1. Not Guilty	5.5	10	10
2. Guilty of causing grievous bodily harm	20	30	32.5
3. Guilty of using more force than necessary	60	50	47.5
4. Hung Jury	10	11.25	10
Jury Case B			
1. Not Guilty	5	5	10
2. Guilty of Causing grievous bodily harm	40	22.5*	27.5
3. Guilty of using more force than necessary	50	54	55+
4. Hung Jury	7.5	10	5

[ ] This is the outcome reported to have occurred. Values with * or + denote that comparison with appropriate Foresight judgement is significantly different *P < 0.05; +P < 0.1.

differences in the expected direction. These were "The British-Gurka Struggle", outcome 2, hindsight after ($P < 0.1$, Mann Whitney $U = 27$). Appendix 6B summarises all the values obtained from foresight comparisons with the two hindsight conditions (HB and HA) using the Mann Whitney U test.

(2) Confidence Levels

Table 6.4 Median Confidence Levels given in Hindsight (HA and HB) and Foresight

SCENARIO	CONDITION		
	FORESIGHT	HINDSIGHT BEFORE	HINDSIGHT AFTER
Mrs. Dewar in Therapy	3.5 N=8	3.0 N=9	4.0 N=9
The British-Gurka Struggle	3.5 N=9	3.0 N=9	2.0 N=9
Jury Case A	4.0 N=9	3.0 N=9	4.0 N=9
Jury Case B	4.0 N=6	4.0 N=8	4.0 N=9

Table 6.4 summarises the median confidence levels made in foresight and the two hindsight conditions. From this Table it can be seen that only one of the eight medians is greater in hindsight than foresight, but not significantly so. Four were less in hindsight and three the same as foresight. Again none of these were significantly different. The table also shows that not everybody gave confidence estimates (for each cell N should equal 10). This was due to a fault in the layout of the questionnaire: the confidence scale appeared on its own on the back page and not everybody noticed it.

Figure 6.1 Median likelihood assessments in foresight and hindsight conditions for the outcome "Guilty of using more force than necessary" (This is the reported outcome).

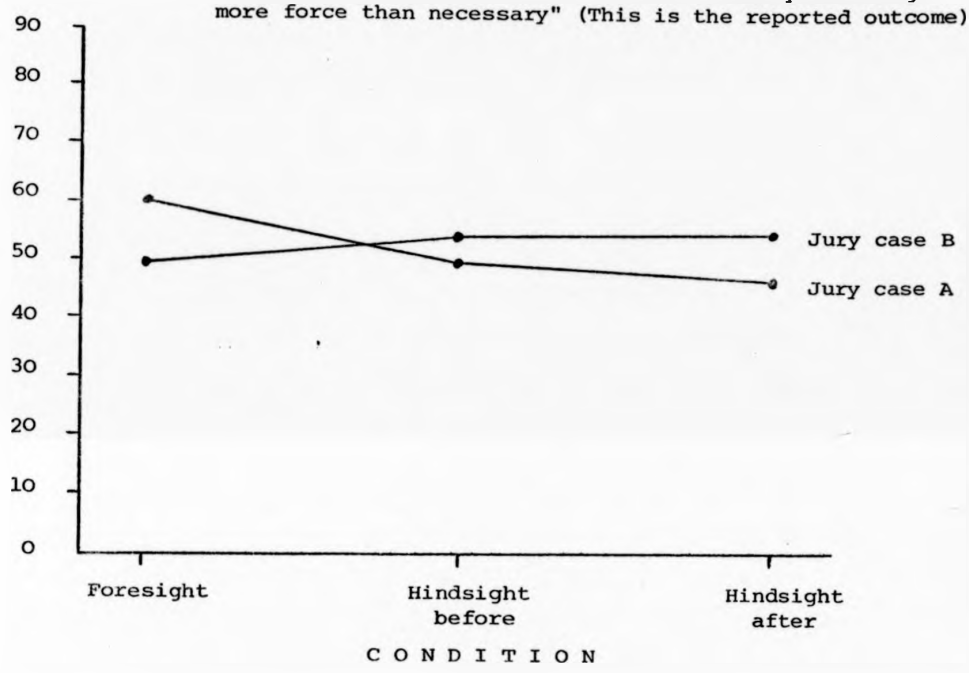
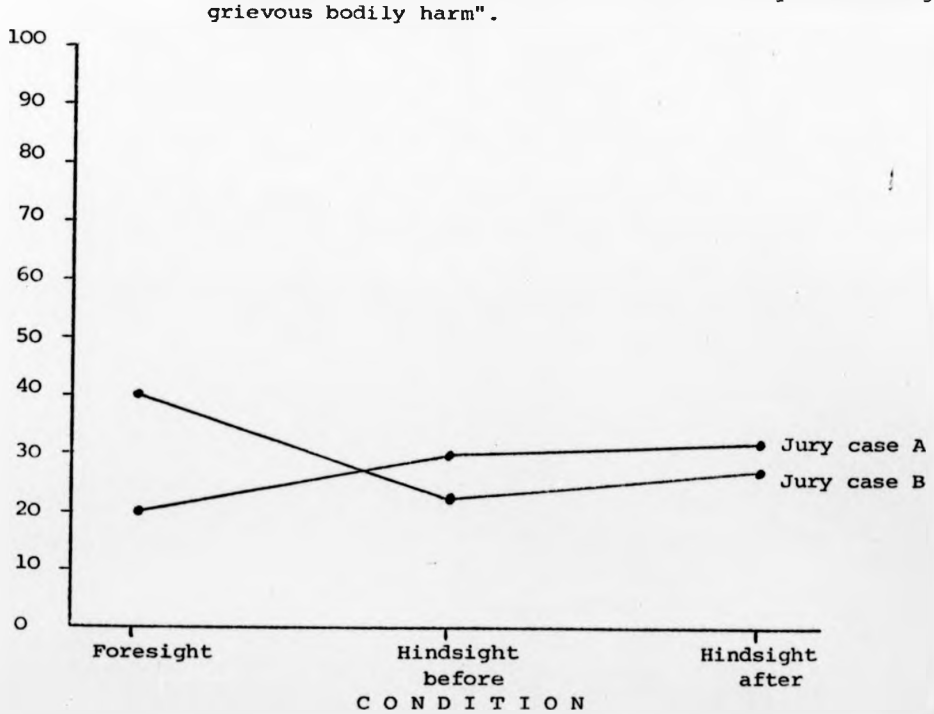


Figure 6.2 Median likelihood assessments in foresight and hindsight conditions for the outcome "Guilty of causing grievous bodily harm".



(3) Creeping Determinism and Information Order in Jury Cases A and B

Figures 6.1 and 6.2 graphically depict the medians for outcomes 2 and 3 respectively in Jury Case A and B. The only difference between foresight and hindsight comparisons, as reported earlier, was a trend in Jury Case B for Outcome 3 between foresight and hindsight after ($P < 0.1$, Mann Whitney U = 29, one tailed). Comparisons between HA and HB for outcomes 2 and 3 ("guilty of causing grievous bodily harm" and "guilty of using more force than necessary" respectively) yielded no significance differences. The only difference between these two information orders was in foresight assessments for outcome 3. Here the assessments given in Jury Case B were significantly higher than those in Jury Case A ($P < 0.05$, Mann Whitney U = 23.5, two tailed).

6.4 DISCUSSION

The first hypothesis this experiment was designed to explore, that creeping determinism would be more in evidence when outcome knowledge was given after the event description rather than before, received only weak support. Where outcome knowledge was given before the event description no evidence of creeping determinism was found. However, two scenarios out of the four produced results consistent with the creeping determinism hypothesis in the condition where outcome knowledge was given at the end of the event description. The pooled results in Tables 6.1 and 6.2 tell a similar story. Overall, less than 50 per cent of subjects or assessments given in hindsight were consistent with creeping determinism. Again only one scenario ("The British-Gurkha Struggle") provided clear evidence for both the outcome reported to have

occurred and the outcomes reported not to have occurred.

In summary, then, differences between foresight and hindsight (after) were much less consistent than those obtained by Fischhoff (1975a). Limited support was obtained for the suggestion from Experiment 1 that creeping determinism is mediated by the position of outcome knowledge. Reporting the outcome of an event produced a statistically significant increase in its perceived likelihood for only two of the four cases and then only when outcome knowledge was presented at the end of the scenario. At the moment the available data would seem to suggest that presenting outcome knowledge at the end of a scenario is a necessary, but not sufficient, condition for hindsight-foresight differences to occur.

The second hypothesis, that subjects would exhibit greater confidence in likelihood assessments made in hindsight than foresight, received no support. This might have been expected in view of the failure to detect consistent effects in hindsight. In the one case where creeping determinism was found ("The British-Gurka Struggle", hindsight after) confidence levels were opposite to that predicted, but not significantly so.

The third hypothesis, concerning the effect of different information orders, manipulated (as indicated in Experiment 1) by ordering statements scaled by Thibaut and Walker, on hindsight judgements also received little support. The hypothesis, it will be recalled, was that hindsight judgements would exhibit stronger creeping determinism when information pointing at the reported outcome was presented last rather

than first. Inspection of figures 6.1 and 6.2 reveals the medians associated with each scenario to be in the predicted direction. For the outcome which occurred (Number 3) assessments are higher, for both hindsight conditions, in Jury Case B. Further inspection of Figure 6.2 reveals the opposite to be the case for the outcome reported not to have occurred (Number 2). As none of these differences were significant, however, the results will not be discussed in any more detail.

However, the interesting question remains as to why quite strong evidence for recency effects in hindsight judgements was found in Experiment 1 but not here. This failure to replicate the results may stem from two considerations. First, it may be due to the procedural differences between the two experiments. In Experiment 1 an item-by-item sequential presentation together with "surprise" instructions in hindsight was used. Experiment 2 presented subjects with a continuous prose passage and instructed subjects to make judgements in the light of all information presented. Both the sequential presentation and "surprise" instructions may account for the failure to replicate order effects here. For example, the sequential presentation procedure may cause people to attend to the information more especially that towards the end. In recalling it in order to formulate likelihood assessments the information which appeared last may be more available from memory and hence exert greater influence over assessments.

The second consideration stems from the way in which information order was manipulated in the two scenarios. In Experiment 1 it was reported as being achieved in a non-rigorous or objective way - the

bulk of "lawful" or "unlawful" statements either appeared first or last, and were ordered in such a way as to make the scenario easy to comprehend.[1] Such a procedure falls short of a rigorous order manipulation in a number of ways. First, rather than have the bulk of lawful or unlawful statements first or last it would be more appropriate a manipulation of order to have all of each sort appear in such a way. Second, the scale values and categorisation given to these statements by Thibaut and Walker's (1975) judges may differ on both counts to those given by British judges. Third, not all the statements taken from Thibaut and Walker were scaled by them, specifically, those briefly outlining the incident. Finally, statements were scaled according to degrees of lawfulness or unlawfulness but not in relation to a specific charge.

In view of these shortcomings a further experiment was designed to investigate the effect of different information orders on hindsight judgements. The ordering of information was achieved using a more rigorous and objective procedure in line with the points discussed above.

---

1 An oversight and shortcoming with this, and the previous experiment, was that subjects were not informed about a specific charge against Adams (the defendant).

CHAPTER 7.

EXPERIMENT 3: THE EFFECT OF INFORMATION ORDER UPON HINDSIGHT JUDGEMENTS

7.1. INTRODUCTION

Experiment 3 is the last one in this thesis to use "laboratory material" to investigate the effect of outcome knowledge upon judgements. The experiment to be reported here returns to an investigation of hindsight bias and how it might be influenced by the order in which information describing some social event appears.

Three hypotheses were investigated with respect to hindsight judgements. That those in possession of outcome knowledge would (1) see that outcome as more likely than subjects who did not know how events turned out; (2) be less surprised at its occurrance than subjects in foresight and, conversely, hindsight subjects would be more surprised if the outcomes reported not to have occurred had occurred than those in foresight; (3) show greater confidence in their judgements than those in foresight.

With respect to information order it was hypothesised that hindsight bias would be more in evidence when information pointing to the outcome believed to have occurred appeared at the end of the event description than at the beginning.

The experiment to be reported here returned to an investigation of hindsight bias (as in Experiment 1) rather than creeping determinism

(as in Experiment 2) for two reasons. (1) as discussed in Chapter 2 it is the most interesting of the two phenomena because of its potentially serious consequences in terms of what one learns from the past. (2) Neither of the previous experiments provide strong support for hindsight bias or creeping determinism. In view of this there appears to be no particular advantage, as argued in the previous chapter, in investigating the latter rather than the former. In consequence the former reason becomes even more compelling.

The three hypotheses concerning the effect of outcome knowledge upon perceptions stem, as in the previous experiments, from Fischhoff (1975a),Fischhoff and Beyth (1975). Hypotheses (1) and (3) need no further introduction as their derivation is discussed in the introductions of Chapters 5 and 6 respectively. Hypothesis (2) derives from Fischhoff (1975) where he argues that people in hindsight will be less surprised at the occurrance of an outcome they believe has occurred than if they had not known how events turned out. In effect, then, hindsight bias should cause people to be less surprised at outcomes believed to have occurred. For outcomes believed not to have occurred people should show greater surprise, as a consequence of hindsight bias, than those in foresight.

The order in which information appeared in the two scenarios used in this experiment ("Mrs. S" and "Adams and Zemp") was varied in two ways. In one order information indicating the outcome reported in hindsight appeared first followed by information indicating other outcomes. In the other this order was reversed. As the hypothesis indicates recency effects are assumed to be operating (if there are

any order effects at all). The reasons for this have been discussed in earlier chapters and will not be repeated here. However, as information order is manipulated in a different way to that in previous experiments a brief summary of the procedure might prove useful.

At the end of the last chapter it was indicated that the way in which information order had been manipulated in the jury cases was not entirely satisfactory or objective. The present study overcame these shortcomings by having subjects scale statements according to Thurstone and Chave's (1929) equal-appearing interval scale. Having obtained a scale value for each statement order of information was varied in one of two ways, as indicated earlier, by rank-ordering the statements according to their value.

In summary, the experiment reported here investigated how subjective likelihood assessments, surprise and confidence levels are affected by possession of outcome knowledge. Secondly, how hindsight bias may be influenced by the order in which statements describing the event appeared.

7.2. METHOD

SUBJECTS

60 undergraduate students at the University of Warwick acted as subjects in this experiment.

DESIGN

The design of the experiment can best be understood by examining Figure 7.1. The study included a foresight and hindsight condition

(where outcome knowledge was given before the event description) with two treatments in each.

Figure 7.1 Design of Experiment 3

FORESIGHT				HINDSIGHT			
ORDER 1		ORDER 2		ORDER 1		ORDER 2	
Mrs. S	Adams & Zemp	Mrs. S	Adams & Zemp	Mrs. S	Adams & Zemp	Mrs. S	Adams & Zemp
N = 15		N = 15		N = 15		N = 15	

In the "order 1" treatment subjects read an event description where the order of information was such that that indicating the reported outcome appeared last and that indicating other outcomes appeared first. In the "order 2" treatment the event description appeared in the reverse order to that above. Subjects in each treatment made two sets of judgements, in a counterbalanced design, one relating to the Mrs. S. scenario and the other to the Adams and Zemp scenario.

PROCEDURE

Thirty subjects were randomly allocated to a foresight condition and thirty to an hindsight condition. In both conditions 15 subjects were randomly allocated to an "order 1" treatment and 15 to an "order 2" treatment.

In both conditions subjects had to make three sets of assessments. They had to (1) assess the likelihood of occurrance of each of the four possible outcomes to the scenario, ensuring that these four likelihoods

totalled 100 per cent; (2) for each outcome indicate, on a seven-point scale ranging from surprised to not surprised, how surprised they would be (have been) had it occurred; (3) for each scenario indicate how confident they felt about the judgements they had just made. Appendix 7A and 7B provide examples of the questionnaire sheets for Mrs. S. in Foresight order 1 (FO1) and hindsight order 1 (HO1) respectively. It should be noted that the questionnaire sheets for each order treatment in each condition were identical.

Subjects in both conditions spent, on average, between 20 and 25 minutes completing the task. Subjects were run in small groups (of between 3 and 6) with the experimenter present to ensure no communication occurred between subjects.

INSTRUCTIONS

In the foresight condition the cover sheet of each questionnaire ran as follows:

> In this study we are interested in knowing how people judge the likelihood of possible outcomes of social events. The relevance of this can be seen by, for example, thinking of some past events that you know about -- how likely was the outcome that actually occurred? How likely did other outcomes seem before the one that actually occurred? Was the outcome that did in fact happen surprising? These are the kind of issues looked at here.
>
> In what follows you will find there are two event descriptions -- one concerned with a woman who has made a number of suicide attempts in the past; the other concerns a summary of evidence heard in a criminal trial. Associated with each event is a short questionnaire.
>
> Would you please proceed as follows:-
>
> Read carefully through the first event description, then answer the questionnaire associated with it. Then go on to the second event description, read carefully through and answer the questionnaire for that one.

Upon reading this and the event description subjects were provided with a questionnaire asking them to make likelihood assessments, estimates of surprise and indicate confidence levels. As an example, the questionnaire for "Mrs. S." is reprinted in Appendix 7A. Appendix 7B is the hindsight questionnaire.

In the hindsight condition subjects received the same instructions on the cover sheet of the questionnaire but with an additional paragraph. This appeared after the second paragraph of the foresight instructions and read as follows:

> As the interest is with past events you will find that before the event description you are told what actually happened. This information is given to you as it provides a realistic simulation of how we view past events. For example, an historian trying to understand why the bombing of Pearl Harbour took America by surprise knows what happened and analyses events in the light of this. Of course, the fact that a certain outcome occurred does not mean it was inevitable, it may have been one that people thought exceedingly unlikely.

These are an elaboration of the "surprise" instructions used in Experiment 1. They implicitly ask subjects to attempt to reconstruct a foresight state of mind. Hence the experiment is investigating hindsight bias.

STIMULUS MATERIAL

Subjects in this experiment were asked to make judgements about two scenarios, one labelled "Mrs. S." and the other "Adams and Zemp". The latter contains exactly the same material as the jury cases in former experiments. The former is new, and was written by the author. It provides a short summary of the personal history of Mrs. S. together

with details about a number of suicide attempts she has made. Appendix 7C reprints this scenario in the Order 1 treatment.

The use of these event descriptions was guided by a number of reasons. First, as the Adams and Zemp scenario had been used before to look at the effect of varying order of information it is of interest to see how previous results compare with those manipulating order in a more objective way. A new event description was used for three main reasons: (1) to ensure that previous findings were not specific to a scenario but would generalise across scenarios. (2) This scenario would draw on less background knowledge, than, for example, the British-Gurkha Struggle. (3) It was felt that statements in this event description would be easier for subjects to scale.

For the Adams and Zemp scenario the possible outcomes listed were the same as in previous experiments, as was the outcome reported to subjects in hindsight (guilty of using more force than necessary). For the scenario about Mrs. S. the outcomes listed were as follows:-

1. Successful suicide attempt
2. Further suicide attempt which was regarded as being highly dangerous to life and of serious intent on her part.
3. Further suicide attempt which was regarded as not being dangerous to life or of serious intent on her part.
4. No further suicide attempt.

For this scenario subjects were asked to make judgements in respect of these outcomes in the 5 year period following Mrs. S's third suicide attempt.

SCALING PROCEDURE AND ORDER MANIPULATION

(a) Scaling

Twenty subjects scaled each statement appearing in both scenarios according to Thurstone and Chave's (1929) equal appearing interval scale. For the scenario "Mrs. S." judges were first asked to read through the event description and scale the statements in accordance with the following instructions:

> Each sentence appears separately on a set of numbered cards. The numbering corresponds to the position of each sentence in the passage that you have just read.
>
> I would like you to describe each of the given sentences using one of the following nine point scales:- One scale concerns sentences which you think indicate that Mrs. S. will not make a further suicide attempt within the five years following her third suicide attempt. The other scale concerns sentences which you think indicate that she will make a further suicide attempt within 5 years following her third suicide attempt.
>
> For the scale concerned with sentences indicating that she will make a further suicide attempt within the next 5 years - on card 9 put those statements which you think strongly indicate no further attempt, on card 1 put those statements which you think weakly indicate no further attempt. Arrange each of the sentences in accordance with the degree of strength or weakness that you think they imply between these two points.
>
> Would you please start with sentence 1 and carry out the above procedure for each of the 16 sentences.

Thus judges had to make two decisions, (1) they had to choose whether the sentence indicated the possibility of a further suicide attempt

or not. (2) They had then to decide, on the appropriate nine-point scale labelled as described in the instructions, to what extent the sentence indicated one of these two things.

Judges scaling the statements of the Adams and Zemp scenario followed a similar procedure but with the following modifications: (1) They had to decide whether the statement implied guilt or innocence for the defendant (Adams) (2) To indicate the degree of guilt or innocence on the appropriate nine-point scale. The two sets of scales for each scenario are reproduced in Appendix 7D.

The scale value of each sentence was determined in accordance with Thurstone and Chave's procedure but with the following modification. For each scenario the two scales (guilty/not guilty: suicide attempt/no suicide attempt) were treated as a continuum from -9 to +9. For Mrs. S. -9 most strongly indicated a further suicide attempt and +9 most strongly indicated no further suicide attempt. For the Adams and Zemp scenario -9 indicated the greatest degree of guilt and +9 the greatest degree of innocence.

The number of times each statement had been given a certain scale value was determined. A cumulative frequency distribution for each statement was computed, starting at the (+) or (-) end depending on which contained over 50% of the scores. A statement's scale value was found by plotting a graph of the cumulative frequency and reading off the 50 per cent level. Thus the scale value is such that half the judges classified a statement to the left of it and half to the right of it. Appendices 7E and 7F summarise the cumulative

frequencies and scale value for each sentence. Tables 7.1 and 7.2 show the sentence, its number, and scale value determined by the above procedure for the scenarios Mrs. S. and Adams and Zemp respectively.

(b) Order Manipulation

The event description information for each scenario was ordered in two ways. In the case of Mrs. S. this was achieved as follows: "Order 1" - the statement with the highest value on the "no further suicide" scale appeared first. This was followed by the statement with the next highest value on this scale, and so on. When all the statements on this scale had been ordered in this manner the statement with the lowest value on the "further suicide attempt" scale appeared next. This was followed by the next highest statement, and so on. The event description ended with the one having the highest value. In "Order 2" statements appeared in the reverse order.

Much the same procedure for the Adams and Zemp procedure was followed. Here the "Order 1" treatment had the statement with the highest value on the "not guilty" scale appear first, and the highest value on the "guilty" scale appear last. Statement in between were ordered as described above. "Order 2" statements appeared in the reverse to that of "Order 1".

7.3 RESULTS

(a) Hindsight Bias

Two sets of analyses were performed to determine the effect of

Table 7.1 Sentence, Scale Value and Sentence Number for Mrs. S. scenario

MRS. S.

Key (-) refers to scale concerned with a further suicide attempt
(+) refers to scale concerned with no further suicide attempt

SCALE VALUE	SENTENCE NUMBER	SENTENCE
-7.7	7	At 27 Mrs. S. made a second suicide attempt by poisoning - this attempt was seen as dangerous to life and of serious intent on her part - she was admitted to a mental hospital for 2m.
-6.0	14	A year after her husband's death she made a third suicide attempt by taking a small overdose of aspirins - this was seen as relatively harmless to life and not of serious intent on her part.
-5.7	11	When she was 40 her husband died of a heart attack which made her depressed and discontented.
-5.0	5	After the first suicide attempt Mrs. S. was restless and discontented perhaps because she did not have a steady boyfriend.
-4.5	2	Mrs. S. was born in 1930 and came from a family in which her father had committed suicide and her mother often threatened it.
-4.2	4	At the age of twenty she made her first suicide attempt by trying to poison herself with her mother's sedatives - this was classified as a dangerous threat to life but not of serious intent on her part.
-3.2	10	Mrs. S. was not altogether happy with her life as a housewife, it often made her tense and depressed
+1.3	16	Mrs. S. had no previous psychiatric history of mental illness apart from that associated with her second suicide attempt.
+2.5	12	Insurances had left her well provided for materially and financially after her husband's death.
+3.3	3	Mrs. S. had had a fairly happy childhood and did not want for any material things.
+4.3	9	In a period of seven years, from 30 to 37, Mrs. S. had three children, two girls and a boy.
+4.7	8	At 29 Mrs. S. was married, she gave up her secretarial work to have a family.

contd....

Table 7.1 contd.

SCALE VALUE	SENTENCE NUMBER	SENTENCE
+5.0	6	She left home at the age of 21 and took a job as a secretary at which she was quite competent and found enjoyable.
+5.0	1	Mrs. S. was a likeable person who always appeared to be cheerful and had a free and easy attitude to life.
+5.0	13	A short while after her husband's death Mrs. S. returned to work as a secretary and seemed to enjoy going back to work.
+7.4	15	Mrs. S. loved her three children and appeared to get a sense of fulfilment from them.

Table 7.2 Sentence, Scale Value and Sentence Number for "Adams and Zemp" scenario

ADAMS AND ZEMP

Key (-) refers to scale concerned with guilt
(+) refers to scale concerned with not guilty

SCALE VALUE	SENTENCE NUMBER	SENTENCE
-8.0	5	Adams responded by stabbing Zemp in the stomach with a piece of glass
-6.4	6	The defendant (Adams) ran out of the back door after the stabbing.
-5.4	10	Adams was not injured when he was knocked to the floor.
-5.4	7	The law provides that it is unlawful to use more force than a person believes necessary in repelling an attack.
-5.3	8	The defendant (Adams) is 6ft. tall and weighs 16 stones, Zemp is 5ft 8" tall and weighs 12 stones.
-4.4	13	One of the waiters told the police he wouldn't mind if Adams never came to the tavern again because he sometimes bullied other customers.
+2.0	2	Recently Adams and Zemp had begun to gamble heavily, and as matters became involved, met at a tavern to discuss their relationship.
+2.0	11	On the night of the stabbing Adams drank only one beer.
+4.3	3	At the start of the meeting Adams told Zemp that they should end their relationship before serious trouble developed between them.
+4.5	14	Zemp told one of the waiters at the tavern he planned to settle his differences with the defendant once and for all.
+5.0	9	Zemp liked to drink heavily and often became violent when he was drunk.
+5.7	4	After a period of conversation Zemp knocked Adams to the floor.
+6.0	1	Adams (the defendant) and Zemp had been close friends for years.
+7.4	12	Before he met Zemp at the tavern the defendant told a friend he hoped he could settle his dispute with Zemp amiably.

outcome knowledge upon likelihood assessments, estimates of surprise and confidence in those judgements. (1) Results were pooled, using the procedure outlined in Chapter 5, to get an overall picture. (2) Individual comparisons between foresight and hindsight judgements were made using the Mann Whitney U test.

Table 7.3 summarises the overall percentages of hypothesis supporting and non-supporting judgements for the outcomes reported and not reported to have happened. As can be seen from this table the results, in general, provide little support for hindsight bias. Only one of the hypothesis-supporting percentages is greater than 50 per cent, this only just so.

Tables 7.4, 7.5 and 7.6 summarises the medians of the likelihood assessments, surprise estimates and confidence levels respectively for the scenarios Mrs. S. and Adams and Zemp, Table 7.4 reveals that of the four hindsight likelihood assessments given to the outcome reported to have occurred two were in the predicted direction and two opposite when compared with respective foresight assessments (FO1 with HO1 and FO2 with HO2). For the scenario Mrs. S. both assessments in the predicted direction were significant (FO1/HO1: $P < 0.01$, Mann Whitney U = 53; FO2/HO2; $P < 0.025$, Mann Whitney U = 56.5). No significant differences between foresight and hindsight assessments for the Adams and Zemp scenario were found. Table 7.4 also shows that for outcomes reported not to have occurred five out of six hindsight assessments were in the predicted direction for the Mrs. S. scenario. Only one of these was significant however (further suicide attempt: not dangerous to life, FO1/HO1, $P < 0.05$,

Table 7.3 Percentage of Hypothesis-Supporting and Non-Supporting Judgements for Outcomes Reported and not Reported to have Occurred

JUDGEMENTS	% SUPPORT	
	(+)	(-)
LIKELIHOOD ASSESSMENTS		
1. Outcomes reported to have occurred	53.3	46.7
2. Outcomes reported not to have occurred	45.5	54.5
SURPRISE ESTIMATES		
1. Outcomes reported to have occurred	46.7	53.3
2. Outcomes reported not to have occurred	48.3	51.7
CONFIDENCE LEVELS	41.7	58.3

Table 7.4 Median Likelihood assessments assigned to the outcomes of the scenarios Mrs. S. and Adams and Zemp

SCENARIO AND OUTCOME	FORESIGHT		HINDSIGHT	
	ORDER 1	ORDER 2	ORDER 1	ORDER 2
MRS. S.				
1. Successful Suicide Attempt	10	10	40**	40*
2. Further Suicide Attempt - Dangerous to Life	25	20	20	20
3. Further Suicide Attempt - not Dangerous to Life	40	40	15***	29
4. No further Suicide Attempt	10	25	5	5
ADAMS AND ZEMP				
1. Not Guilty	10	5	10	5
2. Guilty of Causing Grievous Bodily Harm	20	30	20	40
3. Guilty of using more force than necessary	50	50	40	40
4. Hung Jury	10	10	15	10

[ ] This is the outcome reported to have occurred

Values with * denote that comparison with equivalent foresight condition is significantly different. * $P < 0.025$; $P < 0.01$; $P < 0.005$.

Table 7.5 Median surprise estimates given to the outcomes of the scenarios Mrs. S. and Adams and Zemp

SCENARIO AND OUTCOME	FORESIGHT		HINDSIGHT	
	ORDER 1	ORDER 2	ORDER 1	ORDER2
MRS. S.				
1. Successful Suicide Attempt	4.5	5.0	3.0	4.0
2. Further Suicide Attempt - Dangerous to Life	4.0	4.5	2.5	4.0
3. Further Suicide Attempt - Not Dangerous to Life	2.0	3.0	5.0***	3.0
4. No Further Suicide Attempt	5.5	5.5	6.5	6.0
ADAMS AND ZEMP				
1. Not Guilty	6.0	6.5	6.0	6.0
2. Guilty of Causing Grievous Bodily Harm	4.0	3.5	6.0	4.0
3. Guilty of using more force than necessary	2.0	3.0	3.0	3.0
4. Hung Jury	4.0	6.0	5.0	5.0

[ ] This is the outcome reported to have occurred

Values with * denote that comparison with equivalent foresight condition is significantly different *** $P < 0.001$.

Table 7.6 Median Confidence Levels Given to Judgements made about the Two Scenarios in Foresight and Hindsight

SCENARIO	FORESIGHT		HINDSIGHT	
	ORDER 1	ORDER 2	ORDER 1	ORDER 2
MRS. S.	5.0	5.0	5.0	5.0
ADAMS AND ZEMP	5.0	5.0	5.0	5.0

Mann Whitney U = 50). For the Adams and Zemp scenario all six hindsights assessments were in the opposite direction to that predicted, but none significantly so.

Table 7.5, which summarises the median surprise estimates shows that for the outcomes which occurred two of the four hindsight estimates were in the predicted direction, but none significantly so. For outcomes not reported to have occurred half were in the predicted direction and half in the opposite direction. One of those in the predicted direction was significantly different (Mrs. S., further suicide attempt; not dangerous to life, FO1/FO2; $P < 0.001$, Mann Whitney U = 23).

Table 7.6, summarising the confidence levels, shows all the medians to be identical. Obviously, no significant differences were found here.

(b) Information Order

The lack of strong or reliable support for the notion that oucome knowledge leads to systematic bias means that our predictions concerning the effect of information order on hindsight judgements are unlikely to receive support either.

Inspecting the median likelihood assessments for outcomes reported to have occurred in Table 7.4 reveals that HO1 and HO2 are the same for both scenarios. As expected from this no significant differences were found. For outcomes reported not to have occurred

the table shows only two of the six comparisons to be in the predicted direction. One of these being significant (Mrs. S., further suicide attempt: not dangerous to life, HO1/HO2, $P < 0.025$, Mann Whitney $U = 59.5$). None of the four in the direction opposite to that predicted were found to be significantly different.

The effect of information order upon surprise estimates can be seen from inspection of Table 7.5. For outcomes reported to have occurred it can be seen that Order 1 produces less surprise than Order 2 for the Mrs. S. scenario, however this was not found to be a reliable difference. Similarly, no significant differences were found between HO1 and HO2 for the Adams and Zemp scenario. For outcomes reported not to have occurred only one of the six was in the predicted direction (Mrs. S., further suicide attempt; not dangerous to life), but not significantly so. The five in the opposite direction were not significant either.

7.4 DISCUSSION

The first hypothesis this experiment was designed to investigate, that hindsight bias would be shown in likelihood assessments, estimates of surprise and confidence levels, received only weak support. Likelihood assessments, the traditional way of detecting this bias, provided no support for Fischhoff's claim that possession of outcome knowledge leads people to perceive the outcome they believed has occurred as more likely (and less likely for outcomes believed not to have occurred) than when those same events are seen in foresight when the assessments were pooled together. Only 53.3% of assessments supported this for outcomes believed to have occurred and less than

50% for those believed not to have occurred.

Individual comparisons of foresight and hindsight assessments for each scenario and outcome provided only weak support for hindsight bias. This coming almost entirely from the outcome reported in hindsight in the Mrs. S. scenario. Here hindsight assessments for both order treatments were significantly different from foresight ones. The only other significant result, which was for an outcome believed not to have occurred for the Mrs. S. scenario, also provided evidence of hindsight bias. The failure to detect hindsight bias in the Adams and Zemp scenario is puzzling in view of results in previous experiments where the same scenario provided some support for this bias. Some possible reasons for this will be discussed later.

Comparison of foresight/hindsight estimates of surprise provided virtually no support for the hypothesis that those in possession of outcome knowledge would be less surprised for the outcome believed to have occurred and more surprised for those believed not to have occurred than those in foresight. The pooled results showed slightly more than 50% of estimates to be in the opposite direction. Individual comparisons yielded only one significant difference which was in the predicted direction. Confidence judgements, however analysed, provided no support for the hypothesis.

In summary, foresight/hindsight judgements provided only very weak support for the existence of hindsight bias. This coming almost entirely from subjective likelihood assessments. On the whole, these findings do not provide a convincing pattern of results

for one to claim that possession of outcome knowledge leads to systematic effects.

The second hypothesis concerning the effect of different information orders upon hindsight judgements received little support either. This is partly due, as indicated in the previous section, to little evidence being found for hindsight bias in the first place. No significant differences due to information order were detected for the reported outcomes. For those not reported to have occurred one result was found to be consistent with the hypothesis. This was the only significant finding for all the HO1-HO2 comparisons and constitutes the only support for the second hypothesis.

The lack of a clear pattern of results with respect to both these hypotheses is something of a surprise in view of the findings from the previous two experiments. Although they do not provide convincing evidence for a general hindsight bias, creeping determinism or the effect of different information orders upon hindsight judgements they do provide stronger support than that obtained in this experiment. In view of this we may be justified in speculating upon possible reasons for this discrepancy.

Two considerations may be appropriate here, first, the position of outcome knowledge in relation to the event description; second, the "readability" of the event descriptions. Concerning the former, it will be recalled that outcome knowledge was given before the event description. The previous two experiments provided some evidence, not substantial though, that such position of outcome

knowledge may result in hindsight bias (or creeping determinism) being less pronounced than if it were given after the event description. The design of the experiment reported in this chapter may be criticised on these grounds as controlled variables were designed, or should have been, to encourage hindsight bias.

The second, and more important consideration, concerns the procedure used for manipulating information order. The problem arises from ordering sentences or statements in a scenario with respect to their scale value only. Such an approach ignores the "readability" and overall coherence of the passage. Indeed, many subjects spontaneously commented upon difficulties they encountered when trying to piece together. In the Mrs. S. scenario, subjects found it difficult to sort out the chronological sequence of events. The ordering procedure ignores this aspect, as a consequence the event description jumps from one time period to another. In the Adams and Zemp scenario a similar problem was experienced - this concerned the sequence of events leading up to and following the stabbing. The net effect of the "readability" problem was to cause subjects to read the event description more than once to achieve a clear understanding. This "repeated exposure" to the information, as will be recalled from Chapter 4, is likely to nullify order effects (Anderson and Hubert, 1963).

In summary, the above two considerations indicate that the weak support for the experimental hypotheses may be partially due to shortcomings in design. In particular, the "intuitive" manipulation of information order described in Experiment 1 may be less objective

but does preserve coherence and make understanding relatively easy in a way the more objective procedure does not. Perhaps the lesson to be drawn from this is that some kind of satisfactory and specifiable trade off between these two criteria is required. An easily comprehensible passage is as important as an objective procedure for manipulating orders. It is hard to see how the two could be made completely compatible but some attempt to take account of both factors is required.

The next chapter reports an experiment which continues the main theme of the one reported here. The experiment is concerned solely with the effect of different orders of information, this with respect to jurors' perceptions of guilt or innocence in a rape trial.

CHAPTER 8.

EXPERIMENT 4: EFFECTS OF WITNESS AND TESTIMONY ORDER UPON JUROR VERDICTS

8.1 INTRODUCTION

The experiment to be reported in this chapter differs from the other empirical studies in this thesis as it does not investigate hindsight judgements. The experiment is solely concerned with order effects and investigates this with respect to jurors' verdicts of guilt or innocence using a criminal trial. The reasons for this were summarised in Chapter 1 and discussed at greater length in Chapter 4 which reviewed order effects.

The experiment, whilst preserving the overall structure of adversary proceedings in this country, manipulated the order in which witnesses appeared and the order of testimony given by each witness. Recency effects, it will be recalled from Chapter 4, are the most common findings in psycholegal studies, with Thibaut and Walker (1975) providing the strongest evidence. On the basis of their research the experiment reported here was designed to test two predictions. That (1) there would be a greater number of guilty verdicts when witnesses and testimony indicating guilt appeared last, and, conversely, a greater number of not guilty verdicts when witnesses and testimony indicating innocence appeared last in both the case for defence and prosecution; (2) confidence levels in these judgements would be greater when guilty or not guilty verdicts were given by subjects where witnesses and testimony indicating guilt appeared last

or indicating innocence appeared last respectively.

Thibaut and Walker (1975) make strong claims concerning the importance and relevance of their findings to the courtroom. In commenting upon their results they say: "These results clearly show that in the legal setting recency effects are pervasive" (p.61). To accept such implications of the results of experimental research we need to be confident that such research adequately characterises and simulates criminal courtroom proceedings. Thibaut and Walker's research on order effects, reviewed in some detail in Chapter 4, was found to suffer from a number of serious shortcomings in this respect. The detailed discussion given there will not be repeated here, only a summary of the criticisms will be given. It will be recalled that the most serious shortcomings were (a) poor simulation of criminal courtroom proceedings, (b) inadequate trial material and (c) failure to ask subjects for guilty or not guilty verdicts.

The experiment to be reported here attempts to overcome these problems by: (1) maintaining the traditional procedure in criminal trials (i.e. opening statements by prosecution and defence; the case for the prosecution and defence; closing arguments by prosecution and defence; summing-up by the judge); (2) using a shortened version of an actual (rather than fictional) criminal case - here a rape trial; (3) presenting a version of substantial length (over 3,000 words) rather than a short summary (less than 500 words); (4) asking simulated jurors to give verdicts rather than indications of guilt or innocence on a nine-point scale.

Attempting to provide greater "ecological validity" to the experiment imposes restrictions on the way in which order can be manipulated. For example, one cannot, as most studies have (Wallace and Wilson, 1969; Wilson, 1969), vary the order in which prosecution and defence appear in the proceedings. In view of this, and to provide some similarity between Thibaut and Walkers' "gross" and "internal" order, the present study varied order in two ways (1) Witness order - here witnesses for the defence or prosecution were ordered in terms of degree of guilt or innocence the judges thought they indicated for the defendant. Witnesses called for the defence or prosecution remained in their respective part of the proceedings, only the order in which they appeared was varied. In one condition witnesses indicating strongest degree of guilt appeared first followed by those judged to indicate innocence. The other condition reversed this order (2) Testimony order - here the order of statements made by each witness was varied (with the restriction that defence or prosecution's cross-examination, as appropriate, came last). There were two orders, either strong guilty followed by weak guilty, weak innocent, strong innocent statements; or the reverse of this.

In summary, the present study investigated order effects with respect to witnesses and their testimony. It was predicted that, on the basis of previous research, recency effects would be obtained. In seeking to make the findings applicable and relevant to actual criminal courtroom proceedings the experiment was designed so as to preserve traditional adversary protocol. Previous research has failed to do this, in consequence their findings must be regarded as having little relevance or application for the legal profession

in particular, and the law in general.

8.2. METHOD

SUBJECTS

128 Open University students, 64 male and 64 female, attending a Summer School at Warwick University served as "jurors" in this experiment. People who would be exempt or ineligible to serve on a jury were excluded from the study (see Devlin, 1966, Ch.1. for a detailed list).

DESIGN

Figure 8.1 Design of Experiment 4

SEX	WI		WG	
	TG	TI	TG	TI
MALE	N = 16	N = 16	N = 16	N = 16
FEMALE	N = 16	N = 16	N = 16	N = 16

Key: WI - "innocent" witness first

WG - "guilty" witness first

TG - "innocent" testimony first

WG - "guilty" testimony first

The design of the experiment can best be understood by examining Figure 8.1. The study included two witness order conditions (WG and WI) and two testimony order conditions (TG and TI).[1] In the witness order conditions subjects read a shortened version of a trial transcript in which either "innocent" witnesses appeared first followed by "guilty" ones (labelled WI), or "guilty" appeared first and

---

[1] In the witness order conditions the order in which witnesses for prosecution and defence appeared was varied; in the testimony order conditions the statements made by these witnesses was varied.

"innocent" last (labelled WG). In the testimony order conditions either "guilty" statements followed by "innocent" ones (TG) or "innocent" followed by "guilty" (TI) was presented to simulated jurors. The two order manipulations, therefore, yielded four conditions: WI - TG; WI - TI; WG - TG; WG - TI.

PROCEDURE

Thirty two subjects were randomly assigned to one of the four order conditions, with the restriction that 16 males and 16 females appeared in each.

Each subject received a twelve page booklet, containing instructions, the rape trial (order of witness and testimony according to condition), and a questionnaire. The two page questionnaire is reprinted in Appendix 8A. The instruction sheet, which was on the cover page of the booklet, read as follows:

> Would you please read through the following passage which is a summary of an actual trial. After you have read this summary would you please answer the short questionaire which follows it.

The next ten pages of the booklet were taken up with the rape trial structured, as indicated in the introduction, to conform to traditional British criminal courtroom proceedings. The final two pages of the booklet contained the questionnaire. "Jurors" were asked to make two kinds of judgements: (1) whether they thought each of the two defendants was guilty of rape, attempted rape or not guilty of either charge, (2) To indicate, on a seven-point scale ranging from "not

confident at all" to "very confident", how confident they were about the verdicts they had given.

Subjects spent, on average, 45 minutes reading the rape trial transcript and completing the questionnaire. After completing the task they were debriefed as to the aims of study and invited to make comments about the case, etc.

STIMULUS MATERIAL

The rape trial proceedings were transcribed from an audio tape used by Sealy (1973, 1975) in his experimental studies. The material was derived from an actual trial transcript (Sealy-personal communication). The audio-tape obtained from Phil Sealy lasted for over 2.5 hours and when transcribed amounted to over 12,000 words. In order to reduce this between one third and one quarter two judges (the experimenter being one) were given the task of precising it without changing the structure. A third judge then compared the two versions for discrepancies and/or omissions. Where such problems arose the three judges discussed them and reached consensus over what to include and omit. One of the two versions was selected for use in the experiment.[2]

SCALING AND ORDER MANIPULATIONS

(a) Scaling

Twenty subjects scaled 63 statements of testimony according to

---

2 The two summaries made by the two judges together with the audio-tape transcription can be obtained on request.

Thurstone and Chave's (1929) method of equal-appearing interval scale. A similar procedure to that outlined in the previous experiment (see Chapter 7, pp. 184) was followed. In view of this only the differences will be described here.

Judges were asked to scale the statements in accordance with the following instructions:

> The object of this exercise is to scale evidence from an actual rape trial transcript.
> Would you please read through the prosecution's opening statement, Mary Watkin's evidence and the judge's summing up.
> There are two sets of cards with the following letters on them: A, B, C, D, E, F, I, J, K. Please arrange them before you in that order.
> The evidence from the transcript is numbered, as you will see. For each numbered statement would you carry out the following procedure: for the set of cards relating to Harrison would you put those statements which you think imply the highest degree of guilt on card A. On card F put those which imply a neutral position with respect to Harrison. On card K put those statements which imply the highest degree of innocence for Harrison. On the rest of the cards arrange statements in accordance with the degree of guilt or innocence you believe they imply. Do not try to get the same number in each pile as they are not evenly distributed.
> Would you repeat this procedure for the set of cards relating to Bryce.

From these instructions it can be seen that judges were first required to read through the trial transcript. The order of witness and testimony presented to them was that of the original audio tape transcription as summarized by the judges. Appendix 8B reproduces the trial case that the scaling judges read. As there were two defendants in this case, Harrison and Bryce, judges scaled each statement twice: once on an 11-point scale, ranging from highest degree of guilt to highest degree of innocence, with respect to

Harrison; secondly, on a similar 11-point scale, but this time with respect to Bryce.

The method for determining the scale value of each statement was similar to that outlined in the previous chapter, again only the differences will be described here. First, the scale value of each statement was determined with respect to Harrison. Previous research using this material has indicated that most people perceive Bryce to be innocent, whilst verdicts concerning Harrison tend to be more equivocal. Thus manipulating order with respect to Harrison should stand more chance of detecting order effects, if they exist. Secondly, each statement was classified as indicating "guilt" or "innocence" depending on which side of the mid-point of the scale it fell. Thirdly, a "Q-value" for each statement was determined. This is a measure of ambiguity and is found by computing the scale-span of the two quartile points of the cumulative curve for a statement. Appendix 8C summarises the category, scale and Q-value of each statement together with the cumulative frequencies. The sentence numbers given correspond with the sentence numbers in Appendix 8B.

(b) Order Manipulation

Witness order was varied by first determining the degree of guilt or innocence of each witness as indicated by the twenty judges. This was achieved by selecting the statement or sentence with the highest scale value for guilt or innocence for each witness. This determined the category and ranking of witness. All witnesses for the prosecution fell in the "guilty" category and witnesses for the

defence in the "innocent" category.

Witness order for the WG treatment was achieved simply by ordering the witnesses, for the prosecution's case, from most guilty to least guilty; and, for the defence's case, from least innocent to most innocent. Witness order for the WI treatment was achieved by ordering least guilty to most guilty and most innocent to least innocent for prosecution and defence respectively. This yielded the following witness orders:

WG	WI
Cross-Examination of Mary Watkings	Cross-Examination of Mary Watkins
Detective Constable Foster	George Smith
Dr. Grovernor's Report	Mrs. Godfrey
Dorothy Watkins	Joyce Blunden
Joyce Blunden	Dorothy Watkins
Mrs. Godfrey	Dr. Grovenor's report
George Smith	Detective Constable Foster
Harrison	Bryce
Bryce	Harrison

Testimony order was manipulated by ordering the statements made by each witness in one of two ways: (1) TG - strongest guilty statements, followed by weakest guilty, weakest innocent and strongest innocent. (2) TI - strongest innocent statements, followed by weakest innocent, weakest guilty and strongest guilty.

These two order manipulations yielded four treatments: (1) WG-TG; (2) WG-TI; (3) WI-TG; (4) WI-TI. The testimony orderings are similar to those of "climax" (TI) and "anticlimax" (TG) characterised by Mckelog (1952) and investigated by Thibaut and Walker (1975) under the heading of "internal bias".

8.3. RESULTS

The results are reported in three sections: (1) those relating to verdicts given by individual "jurors" to Harrison and Bryce on the charges of rape and attempted rape (2) Confidence levels given by jurors to those verdicts (3) Sex differences in verdicts.

(1) Verdicts

Tables 8.1 and 8.2 summarise jurors verdicts on the charges of rape and attempted rape for Harrison and Bryce respectively. Table 8.1 shows that, overall, the WG-TG condition produced the greatest number of guilty verdicts (16 out of 32) on the charge of rape for Harrison. The least number of guilty verdicts (6) being obtained in the WI-TG condition. Table 8.2, which summarises the verdicts for Bryce, shows that the number of guilty verdicts on the charges of rape and attempted rape does not differ greatly between conditions.

To discover whether order manipulations affected verdicts four χ^2 were performed, one for each defendant and each charge, for guilty and not guilty verdicts. For Harrison on the charge of rape and attempted rape no differences due to order were found ($\chi^2 = 4.75$, $df = 3$, $P > 0.05$; $\chi^2 = 1.4$, $df = 3$, $P > 0.05$, respectively). Similarly, no reliable differences due to order were found for Bryce on the charge of rape and attempted rape ($\chi^2 = 0.43$, $df = 3$, $P > 0.05$; $\chi^2 = 1.48$, $df = 3$, $P > 0.05$, respectively). Performing a 2×2 χ^2 on testimony and witness order again revealed no significant differences ($\chi^2 = 0.2$, $df = 1$, $P > 0.05$; $\chi^2 = 0.61$, $df = 1$, $P > 0.05$, respectively).

Table 8.1 Summary of Verdicts for Harrison on the Charge of Rape and Attempted Rape

SEX	WG - TG			WG - TI			WI - TG			WI - TI		
	RAPE	ATTEMPT RAPE	NOT GUILTY	RAPE	ATTEMPT RAPE	NOT GUILTY	RAPE	ATTEMPT RAPE	NOT GUILTY	RAPE	ATTEMPT RAPE	NOT GUILTY
MALE	6	4	6	4	6	6	4	6	6	5	6	5
FEMALE	7	7	2	3	8	5	2	12	2	5	6	5
TOTAL	13	11	8	7	14	11	6	18	8	10	12	10

Table 8.2 Summary of Verdicts for Bryce on the Charge of Rape and Attempted Rape

SEX	WG - TG			WG - TI			WI - TG			WI - TI		
	RAPE	ATTEMPT RAPE	NOT GUILTY	RAPE	ATTEMPT RAPE	NOT GUILTY	RAPE	ATTEMPT RAPE	NOT GUILTY	RAPE	ATTEMPT RAPE	NOT GUILTY
MALE	1	2	13	1	4	11	2	2	12	2	2	12
FEMALE	2	5	9	1	0	15	0	3	13	1	2	13
TOTAL	3	7	22	2	4	26	2	5	25	3	4	25

(2) Confidence in Verdicts

Table 8.3 summarises the means and standard deviations for each of the order treatments for Harrison on the charges of rape and attempted rape. To discover if witness and testimony order affected confidence in verdicts by males and females a 3-way analysis of variance was performed. The factors order, verdict and sex were extracted giving a 4 x 2 x 2 analysis with independent measures (Winer, 1971, p.452). Table 8.4 summarises the results of this analysis.

Table 8.3 Summary of Means and Standard Deviations of Confidence Levels for Harrison on the charge of rape and attempted rape

HARRISON - CHARGE OF RAPE									
SEX		WG - TG		WG - TI		WI - TG		WI - TI	
		GUILTY	NOT GUILTY	GUILTY	NOT GUILTY	GUILTY	NOT GUILTY	GULITY	NOT GUILTY
MALE	$\bar{X}$	5.08	4.69	6.13	3.83	5.38	4.42	4.00	4.73
	SD	1.11	2.45	0.25	1.54	0.75	2.01	2.00	1.37
	N	6	10	4	12	4	12	5	11
FEMALE	$\bar{X}$	5.00	4.22	6.30	4.69	5.00	4.11	4.20	4.00
	SD	0.82	1.64	1.15	1.30	1.41	1.62	1.48	1.79
	N	7	9	3	13	2	14	5	11
HARRISON - CHARGE OF ATTEMPTED RAPE									
MALE	$\bar{X}$	4.63	2.83	5.00	3.91	4.42	5.25	3.83	4.00
	SD	1.89	2.48	0.71	1.50	1.63	1.25	1.91	1.87
	N	4	6	6	6	6	6	6	5
FEMALE	$\bar{X}$	5.57	1.50	4.75	3.90	4.63	2.50	5.83	2.80
	SD	0.98	0.71	1.56	1.35	1.43	0.71	0.75	1.92
	N	7	2	8	5	12	2	6	5

Table 8.4 Summary of Analysis of Variance on Confidence Levels of Juror verdicts with respect to Harrison

HARRISON ON THE CHARGE OF RAPE					
SOURCE	SS	DF	MS	F	SIGN
A(ORDER)	12.497	3	4.16	1.59	NS
B(SEX)	0.156	1	0.156	0.06	NS
C(VERDICT)	15.207	1	15.207	5.82	P < 0.025
A x B	2.789	3	0.93	0.36	NS
A x C	15.295	3	5.10	1.95	NS
B x C	0.112	1	0.11	0.04	NS
A x B x C	1.999	3	0.66	0.25	NS
ERROR	292.75	112	2.61		
HARRISON ON THE CHARGE OF ATTEMPTED RAPE					
A(ORDER)	5.49	3	1.83	0.77	NS
B(SEX)	1.70	1	1.70	0.72	NS
C(VERDICT)	42.44	1	42.44	17.94	P < 0.001
A x B	6.95	3	2.32	0.98	NS
A x C	12.48	3	4.16	1.76	NS
B x C	19.99	1	19.99	8.45	P < 0.005
A x B x C	10.57	3	3.52	1.49	NS
ERROR	179.76	76	2.35		

Data for Harrison

No significant differences in confidence levels due to order ($F = 1.59$; $df = 3.112$; $P > 0.05$) or sex of juror ($F = 0.81$; $df = 1.112$; $P > 0.05$ were found. However, a significant difference in confidence levels between verdicts was found ($F = 5.82$; $df = 1.112$; $P < 0.025$). Jurors expressed more confidence in guilty verdicts than not guilty ones. There were no significant interactions. For Harrison on the charge of attempted rape neither order ($F = 0.77$; $df = 3.76$; $P > 0.05$) or sex ($F = 0.72$; $df = 1.76$; $P > 0.05$) produced any reliable differences in confidence levels. Once again, though, jurors expressed greater confidence in guilty verdicts than not guilty verdicts ($F = 17.94$; $df = 1.76$; $P < 0.001$). A significant interaction between sex and verdict was found ($F = 8.45$; $df = 1.76$; $P < 0.005$). No other

significant interactions were detected.

Data for Bryce

The confidence jurors expressed about their verdicts in the different order treatments by males and females for Bryce on the charge of rape and attempted rape are summarised in Table 8.5. A two-way analysis of variance was performed and is summarised in Table 8.6. Extracting order and sex yielded a 4 x 2 analysis with independent measures. The factor guilty/not guilty was not extracted because insufficient numbers of jurors gave guilty verdicts to enable an anlysis to be carried out (Table 8.5 shows this).

Table 8.5 Summary of Means and Standard Deviations of Confidence Levels for Bryce on the Charge of Rape and Attempted Rape

BRYCE ON THE CHARGE OF RAPE								
	WG - TG		WG - TI		WI - TG		WI - TI	
	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE
$\bar{X}$	4.97	4.06	5.25	5.06	5.03	4.66	4.59	3.88
SD	1.70	1.75	1.60	1.12	1.78	2.16	1.81	1.96
N	16	16	16	16	16	16	16	16
BRYCE ON THE CHARGE OF ATTEMPTED RAPE								
$\bar{X}$	4.70	3.93	5.17	4.70	4.80	4.28	4.53	3.80
SD	1.61	1.39	1.19	1.45	1.90	1.77	1.85	2.14
N	15	14	15	15	15	16	15	15

For Bryce on both the charges of rape and attempted rape no signficant differences due to order were found ($F = 1.55$; $df = 3.120$; $P > 0.05$; $F = 1.08$, $df = 3.112$, $P > 0.05$ respectively). The sex of juror produced a trend, showing females to be less confident about their verdicts than males for both charges ($F = 2.89$, $df = 1.120$,

Table 8:6 Summary of Analysis of Variance on Confidence Levels of Jurors with respect to Bryce

BRYCE ON THE CHARGE OF RAPE					
SOURCE	SS	DF	MS	F	SIGN
A(ORDER)	15.3	3	5.1	1.55	NS
B(SEX)	9.5	1	9.5	2.89	P < 0.1
A x B	1.84	3	0.61	0.19	NS
ERROR	395.7	120	3.30		
BRYCE ON THE CHARGE OF ATTEMPTED RAPE					
A(ORDER)	9.93	3	3.31	1.08	NS
B(SEX)	11.42	1	11.42	3.73	P < 0.1
A x B	0.62	3	0.21	0.07	NS
ERROR	342.53	112	3.06		

P < 0.1; F = 3.73, df = 1.112, P < 0.1, respectively). No significant interactions were detected.

(c) Sex Differences

Apart from the above sex differences for confidence levels in juror's verdicts by Bryce, the verdicts themselves were analysed from this perspective. Tables 8.1 and 8.2 summarise verdicts produced by males and females in the different experimental conditions for Harrison and Bryce respectively. Table 8.1 shows that males produced slightly more guilty verdicts on the charge of rape than females (19 to 17 respectively) but females gave a much greater number of guilty verdicts on the charge of attempted rape than males (33 to 22 respectively). Table 8.2 shows males to give a slightly greater number of guilty verdicts for Bryce on the charge of rape than females (6 to 4 respectively). However, the number of guilty verdicts on the charge of attempted rape given by males and females was the same (10 each). Overall, though, most people found Bryce innocent

of both charges.

To discover if there were any significant differences sex was taken out as a factor for the three possible verdicts for Harrison and Bryce. Performing a χ^2 (4 x 2) yielded no significant differences. Pooling male and female responses across the order treatments yielded a 2 x 2 contingency table (guilty/not guilty and male/female). For Harrison on the charge of rape no significant difference was found ($\chi^2 = 0.16$, df = 1, $P > 0.05$), however, on the charge of attempted rape women were found to make more guilty verdicts than men ($\chi^2 = 4.34$, df = 1, $P < 0.05$). No significant differences in verdicts for Bryce were found on either of the charges.

8.4 DISCUSSION

This experiment was designed to explore the effects of different witness and testimony order upon jurors verdicts and their confidence in those verdicts. Neither hypothesis in respect to these judgements received any support. This was surprising in view of Thibaut and Walker's (1975) findings and their claims of powerful recency effects in courtroom proceedings. In view of this some discussion and speculation over why this occurred is warranted.

Four issues need to be considered in this respect, they concern: (a) the complexity and length of the trial case; (b) the simulation of courtroom proceedings; (c) the manipulation of witness and testimony order; and (d) sample size. The first two have implications for the relevance of previous research in this case; the last two are methodological/design views.

In the introduction to this study it was indicated that the experiment reported here differed from previous ones investigating order effects because the stimulus material was of considerable length (over 3,000 words) relative to other studies (around 500 words at most). One obvious explanation for the failure to find empirical support for the hypotheses is simply that order effects disappear with longer and more complex material. Little previous research on order effects in psycho-legal studies specifically and order effects generally appears to have addressed this question.

In Bayesian type judgements a large number of trials has been found to encourage primary effects (Peterson and Duchorme 1967; Roby 1967; Dale, 1968; Miller, 1968). However, because this can be regarded as a form of "conservatism" (see Chapter 4 for discussion of this) and because the stimulus material and task is so different to that used in this study the relevance of such studies to this issue is limited. In view of the lack of empirical enquiry into order effects and length of material the research reported here might well be taken as an indication that long and complex material vitiates such manipulations.

The problem then arises of generalising to trials which last for a number of days or even weeks, which is often the case. Might we expect order effects to reappear under such circumstances? It is not possible on the basis of this and previous research to provide any answer to this. The important point to be made here is that before we can accept the claims of Thibaut and Walker, and other researchers, some empirical work addressing itself to this issue is of vital importance.

A further explanation for the lack of order effects may be due to the way in which this experiment simulated courtroom proceedings. Again, as indicated in the introduction and Chapter 4, this feature is often absent in previous research, in particular cross-examination of the witness is usually omitted. Generally speaking, much previous research on order effects has varied prosecution/defence order. Whilst such an approach might shed light on important theoretical questions over which may be the most fair it has very little to offer in terms of how bias may enter actual practice. Manipulating witness and testimony order is one of the things that can be done within procedural constraints. Lack of any support for such variables affecting verdicts and confidence levels, but especially the former, may imply that they are not important ones to study in this context.

A distinction, however, needs to be made between "climax" and "anticlimax" ordering (McKelvy, 1952; Busch, 1961) and the testimony order manipulated here. Presentation of legal material from weak to strong (as in climax order) may not be the same as the TI order in this experiment. There may be, and probably is, an important difference between an objective and scientific approach, as used here, and one where counsel for defence or prosecution intuitively decides what a "climax" (or "anticlimax") order would consist of. The method for ordering testimony ignores the "gestalt" of a prosecution or defence case. The objective procedure may not result in simulated jurors forming a clear picture of the case, or result in the case having the greatest impact on them. The former view was discussed in the previous chapter and will not be repeated here. The latter problem is one that might be resolved by using a scaling procedure designed

to take account of the internal relationship of testimony or the relation between witnesses.

In summary, the above discussion has argued two things: (1) that the order manipulations used here, within the constraints of traditional criminal proceedings, may not be important variables to study. (2) It may well be the case that "climax" or "anticlimax" ordering is an important variable to study, but the procedure used here and by Thibaut and Walker (1975) does not address it in the sense outlined above.

A number of methodological points with respect to whether a strong enough manipulation of order was achieved in this study, in addition to the discussion above, remain. These may also offer partial explanation for the lack of empirical support for the experimental hypotheses. (1) Only order of witness and testimony was varied, prosecutions' opening statement, summaries by prosecution, defence and the judge were left unchanged from the original transcript. It is possible that manipulation of these along "climax" or "anticlimax" principles may have influenced verdicts. However, the scaling procedure used in this experiment would not be able to achieve this for reasons discussed earlier. (2) The method for scaling witnesses may not have been adequate. A better approach, perhaps would have been to ask judges to scale each witness as well as their constituent statements.

The final consideation bearing on the failure to detect order effects is concerned with sample size. Thirty two subjects were

included in each order treatment, however because three different verdicts were possible with respect to each defendant this sometimes led to small numbers of guilty verdicts on charges of rape and attempted rape. Hence, failure to detect differences due to witness and testimony order may, in part, reflect the small number of subjects who gave such verdicts. A similar experiment using twice as many subjects, for example, would be necessary before we could have confidence in dismissing the idea that such order manipulation might influence verdicts.

The experiment did find differences between males and females in verdicts and confidence in those verdicts. Sex differences for verdicts were evidenced by the two interactions in confidence levels with respect to Harrison. Both interactions indicated that females were more confident of guilty verdicts than not guilty verdicts, whereas there was no difference for males. Both male and female jurors, however, expressed more confidehce in guilty verdicts than not guilty ones. As these findings are not of central importance to the experiment little further discussion will be given to them. It is probably worth noting that experiments using "rape trial" material often report a tendency for females to convict more often than males. As regards the general differences in confidence levels for guilty and not guilty verdicts it appears that once people have made up their mind that a person is guilty they are less likely to change their mind or be persuaded that the person is not guilty. This may be seen as being consistent with the "beyond reasonable doubt" instructions judges give to jurors.

In conclusion, this experiment does not provide evidence for order effects in criminal material using courtroom proceedings. In

view of this the claims of Thibaut and Walker, outlined earlier, must be treated with extreme caution. Further research of this type is required to ascertain exactly what effects, if any, witness, testimony, etc. orders have on verdicts. In general, the results of this and the previous experiment indicate that order effects are much less pervasive than some of the literature would have us believe.

view of this the claims of Thibaut and Walker, outlined earlier, must be treated with extreme caution. Further research of this type is required to ascertain exactly what effects, if any, witness, testimony, etc. orders have on verdicts. In general, the results of this and the previous experiment indicate that order effects are much less pervasive than some of the literature would have us believe.

CHAPTER 9.

EXPERIMENT 5: THE BRITISH FIREMAN"S STRIKE OF 1977/78: AN INVESTIGATION OF JUDGEMENTS IN FORESIGHT AND HINDSIGHT[1]

9.1. INTRODUCTION

The next three experiments to be reported in this thesis all use contemporary socio-political or social events to investigate the effect of outcome knowledge on judgements under uncertainty. The rationale for the importance of an applied emphasis on this type of research was outlined in Chapter 3 and will not be repeated here. Experiment 5, reported in this chapter, investigates hindsight judgement with respect to the British Fireman's Strike of 1977/78. The second experiment, reported in Chapter 10, again investigates the effect of outcome knowledge upon judgements but this time with respect to the outcome of a pregnancy test. The third experiment in this series, reported in Chapter 11, investigates the same thing but with respect to the British General Election of May 1979.

BACKGROUND TO THE FIREMEN'S STRIKE

The Fire Brigade Union (F.B.U.) began a nationwide strike on Monday, 14th November 1977 which ended on Monday 16th January 1978. The strike, which lasted for nine weeks, was over wages. The F.B.U. was asking for a 30% rise; the local authorities offered 10%, in

---

1. A version of this chapter is to appear as an experimental report in the British Journal of Social and Clinical Psychology.

line with the Government (Labour) pay guidelines operating at that time. After much negotiation, with the F.B.U. turning down a number of offers, they eventually settled for 10% with a guarantee that parity with skilled industrial workers would be achieved by November, 1979. The strike was used to investigate the circumstances under which outcome knowledge (hindsight) produces an inappropriate "knew it all along" attitude and to elucidate upon the cognitive processes that may be involved.

Hindsight Bias and the Firemen's Strike

The experiment to be reported here examined three issues with respect to hindsight bias. There were concerned with attempting to determine whether the bias is (1) present in judgements about socio-political events as Fischhoff and Beyth (1975) report; (2) influenced by the way in which likelihood assessments are elicited, and (3) affected by the type of information the judge is given?

Most previous research by Fischhoff and his colleagues has asked judges to assign subjective likelihood assessments to outcomes specified by the experimenter (e.g. Fischhoff, 1975a; Fischhoff and Beyth, 1975). The present study incorporated both this standard technique as well as a new procedure where judges were asked to generate their own outcomes and then assess the likelihood of each. The latter strategy may be seen as a way of aiding judges to recapture or reconstruct a foresight state of mind. Because they have actively thought about the issue and formulated their own outcomes recollection is likely to be more accurate and so hindsight bias is in turn likely to be reduced or eliminated.

The experiment also looked at the effect upon perceptions and judgements of two types of information - a summary of how the strike ended; and details about the first four weeks of the strike. The summary might be expected to encourage hindsight bias as judges are being provided with potential causal explanations of how the outcome that happened came about. If this is so, the outcome that happened should appear more probable: as Carroll (1978) has shown that imagining an event could turn out in a certain way encourages people to perceive that outcome as more probable. Pre-outcome details might be expected to reduce the bias as judges are being provided with ways in which other outcomes could have occurred.

The Effect of Attitudes on Judgement Under Uncertainty

Little research appears to have looked at how attitudes affect judgement under uncertainty, the question has not been addressed at all with respect to hindsight. McGregor (1938), Cantril (1938) investigated various determinants in the prediction of social events. McGregor's interest was in how attitudes, wishes and beliefs together with information influence people when they make (foresight) predictions. In general he found that people who thought, believed or wanted a particular outcome would see it as more likely than those who thought, believed or wanted it not to happen. Recent research has confirmed this with respect to thought (Carroll, 1978).

In view of this it might be expected that judge's attitudes to the Government Pay Policy operating at the time of the Firemen's Strike and attitudes to the Fire Brigade Union's 30% wage claim will affect likelihood assessments assigned to outcomes. We might predict that those who strongly support the Government Pay Policy and oppose

the F.B.U.'s wage claim will perceive outcomes favouring the Government as more likely than those who support the F.B.U. and oppose the Government. Conversely, those for the F.B.U. and against the Government should see those outcomes favouring the F.B.U. as more likely than those for the Government and against the F.B.U.

In summary, using the Firemen's Strike of 1977/78 the effects of outcome knowledge, type of information, strategies of eliciting subjective likelihood assessments and attitudes upon people's perceptions and judgements were investigated.

9.2. METHOD

SUBJECTS

90 subjects, undergraduates at the University of Warwick took part in the experiment.

DESIGN

The design of the experiment can best be understood by examining Figure 1. The study included three conditions (labelled I, II and III) with two treatments in each. In one treatment, "specified", subjects were asked to assess the likelihood of a number of specified outcomes; in the other, "generate", subjects were asked to generate their own outcomes and assign probability estimates to each. In the latter condition subjects were asked to list a series of outcomes, drawing on the "pre-outcome details, which they had just read, and their own knowledge of the strike. They were asked to generate outcomes ranging from unlikely to likely. Subjects were not asked to generate a number of outcomes, but were asked to produce as many as they could think of. Subjects in both treatments were asked to

Figure 9.1 Design of Experiment 5

	CONDITION I		CONDITION II		CONDITION III	
	SPECIFIED	GENERATE	SPECIFIED	GENERATE	SPECIFIED	GENERATE
FORESIGHT	DETAILS PREDICTIONS	DETAILS PREDICTIONS				
HINDSIGHT 6 WEEKS	SUMMARY POSTDICTIONS	SUMMARY POSTDICTIONS	DETAILS SUMMARY POSTDICTIONS	DETAILS SUMMARY POSTDICTIONS	(i) SUMMARY POSTDICTIONS (ii) DETAILS POSTDICTIONS	(i) SUMMARY POSTDICTIONS (ii) DETAILS POSTDICTIONS
HINDSIGHT 3 MONTHS	SUMMARY POSTDICTIONS	SUMMARY POSTDICTIONS				
	N = 15	N = 15	N = 15	N = 15	N = 15	N = 15

TOTAL N = 90

express their likelihood assessments in percentages.

The "specified" outcomes conerned four aspects of the strike: (a) what decision the F.B.U. would reach; (b) how long from Week 4 the strike would continue; (c) what action, if any, the Part Time Fire Fighters would take; (d) what action, if any, the Fire Officers would take with respect to their own separate wage claim. Each of these questions had four or five possible outcomes listed, these are reproduced in Appendix 9A.

"Detail" information consisted of early events, developments, etc. that took place up to and including Week 4 of the strike. The information presented to subjects ran as follows:

Pre-strike Events - Monday 7th November - Saturday 12th November.

> Firemen vote for nationwide strike from 9 p.m. on Monday 14th November by 25,000 to 13,000. This will be the first stoppage in the history of the union (70 years). Current pay for firemen ranges from £50-£65 per week. An offer of 10% has been made by the local authorities in line with the government guidelines, but has been rejected. Troops are being trained to fight fires if the strikes comes off. Government offer of pay deal and reduced hours was rejected by firemen. Also productivity deal was rejected.

Week 1 of the Strike - Monday 14th November - Sunday 20th November

> At 9 a.m. on Monday the firemen implemented their nationwide strike threat. Troops take responsibility for fighting fires. Fire Brigade Union (FBU) claims that response to strike has been solid, the firemen are not getting any strike pay. No early end to strike is foreseeable and negotiations have made no progress. The TUC has not become involved in the dispute so far. Fire officers vote not to go on strike in support of the firemen.

Week 2 of the Strike - Monday 21st November - Sunday 27th November

> No official talks have gone on this week. Part-time firemen who are not on strike have asked for use of fire fighting equipment - the FBU say that to grant this would exacerbate

the situation. The Prime Minister made an appeal to all workers not to break government guidelines, arguing that the best interests of the people of this country are at heart. He further agreed that to give in to the Firemen would mean that everyone would want to break the 10% guideline. It seemed inevitable that the strike would go into third week as Government had made no move to break the deadlock.

Week 3 of the Strike - Monday 28th November - Sunday 4th December

The Government is faced by further challenges from Miners, Civil Servants, and local government workers over 10% guideline. Executive of FBU spoke with the Prime Minister on Monday. Government offer a new pay formula that guarantees that firemen would be exempted from any future pay restrictions. This was rejected out of hand by the FBU, as they insisted on more money now not later. Union was angry, disappointed and determined to strengthen the strike.

FBU asked TUC to start campaign against 10% guidelines being imposed on unions. The TUC replied that they would not fight the government over its pay policy, and thus could not campaign with the Firemen. TUC recommended FBU to negotiate along lines of pay increase and reduction in hours.

FBU say there is no basis on which to negotiate, firemen even more solid than at beginning of strike.

Week 4 of the Strike - Monday 5th December to the present

An opinion poll shows public support for the firemen's strike is waning. Polls shows people believe firemen should settle for 10% rise and reduction in hours. Poll also shows much public support for keeping wage claims to within 10%.

Fire officers have rejected a strike call in pursuit of a 40% claim. FBU say the firemen are as determined as ever, if not more so, as a result of hearing TUC and Fire Officers will not lend support. Pressure exerted by FBU to take action against part-time firemen who are working normally.

The Government have offered a 10% wage rise now with a guarantee that in November '78 and '79 the wages will be reviewed. This is an improvement on the offer made about two weeks ago.

The FBU is to meet to discuss this proposal.

"Summary" information consisted of a 150 word summary of how the strike ended, outlining the meeting that took place, voting for accepting the offer, when the FBU returned to work etc. This ran as follows:

Week 8 and 9 of the Strike - January 2nd to January 13th

On January the 3rd at the request of the Fire Brigades Union a meeting between the Government, local authority employers and the F.B.U. was held. This meeting lasted for three days after which the F.B.U. recalled the delegate conference to decide whether to end the eight-week old strike.

Local authority employers offered a 10 per cent immediately with a guarantee that parity with skilled industrial workers would be achieved by Nov, 1979. This means that the firemen's new basic wage will be £76 on average, and should average £100 by 1979.

On January the 12th the delegate conference voted by a 5 to 2 majority to end the nine-week old strike and accept the 10 per cent now with guarantee of wage rises in Nov. '78 and Nov. '79, whatever pay policy is in force.

On Monday January 16th the firemen returned to work thus ending the first stoppage in the union's history.

Information for both the "summary" and "details" was taken from articles appearing in the "Guardian" and "Times" newspapers.

The attitudes of subjects to the Government Pay Policy and the Fire Brigade Union's 30% wage claim was measured on two seven-point scales. At one end of each scale was written "oppose" at the other end "support". Subjects were asked to circle the number which best reflected their feelings about each of these views. Appendix 9A reproduces these seven-point scales.

PROCEDURE

In Condition I, 30 subjects, 15 randomly allocated to each treatment, made foresight assessments during the fourth week of the strike after first reading the "details". The same subjects made two hindsight assessments, one six weeks and one three months after the strike had ended. These were made after first reading the "summary".

Subjects in both treatments were asked to try to remember or reconstruct the assessments they had given before the outcome was known. Those who had received "specified" outcomes in foresight were given the same outcomes with order and tense changed. In the "generate" treatment the subject was given the outcomes he/she himself/herself had produced in foresight, but without the original probabilities.

In Condition II, 30 subjects, 15 randomly allocated to each treatment, make only hindsight assessments six weeks after the strike had ended. Subjects were given instructions to think back to the middle of the strike (just before Christmas) and, in the "specified" treatment, to assign probabilities they thought they would have given then to the various possible outcomes. (These instructions are similar to those of Fischhoff and Beyth (1975)). In the other treatment, the latter part of the above instructions was modified to include the "generate" procedure. (The actual instructions given to subjects in this condition and the other two are reproduced in Appendix 9B). Subjects then read the "details" and "summary" and then make likelihood assessments as appropriate.

Condition III, again run six weeks after the strike had ended, differed from Condition II in that subjects had to make two sets of assessments. The first was made after they had read the summary (Condition III(i) in Figure 1) the second after reading the details (Condition III(ii) in Figure 1). Again, as in Condition II, 15 subjects were randomly allocated to each treatment.

9.3. RESULTS

To discover how outcome knowledge, strategy of eliciting assessments, type of information and attitudes affected people's judgements, comparisons were made using non-parametric tests (Wilcoxon and Mann Whitney U as appropriate - Siegel, 1956). One-tailed tests were used except where specified.

(1) Data from Treatments when Outcomes Specified

(a) Judgements about the F.B.U. Decision

The analysis first examined judgements about what decision the F.B.U. would reach. In general, the "specified" outcome that actually occurred (F.B.U. get 10% rise and guarantee of increases in November 1978 and 1979) was perceived as more likely in hindsight than foresight. Table 9.1 summarises the median probability assessments and shows that in four out of the five cases assessments in hindsight were greater than in foresight. One of these was significant at the .05 level (Condition II, $U = 70$), another a trend (Condition I, hindsight 6 weeks, $T = 11.5$, $N = 10$, $P < 0.1$.

From Table 9.1 it can also be seen that, for the four outcomes which did not occur, hindsight assessments were less likely only in Conditions II and III, where seven out of the twelve comparisons with foresight in Condition I were in this direction.

Three of these, all in Condition III(ii), were reliable differences - Firemen get 10% and reduced hours, $U = 53$, $P < 0.01$; Firemen get 10% and reduced hours and productivity deal, $U = 59$, $P < 0.025$; compromise somewhere between 10%-20%, $U = 70.5$, $P < 0.05$. None of the

remaining five comparisons was in the opposite direction to that predicted.

For the outcomes which did not happen in Condition I (repeated measures) six were in the predicted direction (4 in hindsight, 6 weeks; two in hindsight, 3 months), and two in the opposite direction. None of these were reliable differences however.

(b) Judgements about the length of the strike

The analysis next examined judgements about how long the strike would continue. Table 9.2 summarises the median assessments and shows that for the outcome which occurred (more than 4 weeks from the fourth week) all five hindsight assessments were significantly more likely than foresight assessments. (Condition I: hindsight 6 weeks, $T = 13$, $N = 13$, $P < 0.01$; hindsight 3 months, $T = 10$, $N = 12$, $P < 0.01$; Condition II, $U = 71$, $P < 0.05$; Condition III:(i) $U = 54.5$, $P < 0.01$; (ii) $U = 54.5$, $P < 0.025$.)

Table 9.2 also shows that for the outcomes which did not occur only the comparisons of Conditions II and III with I (foresight) yielded consistent findings. Seven out of the nine were less likely in hindsight, four being significant. These were: Condition II, from 1-2 weeks longer, $U = 67$, $P < 0.05$; less than a week, $U = 58$, $P < 0.025$; Condition II, less than a week: (i) $U = 71.5$, $P < 0.05$; (ii) $U = 70.5$, $P < 0.05$. Repeated measures comparisons of hindsight/foresight in Condition I for outcomes which did not occur yielded no significant effects.

Table 9.1 Median Likelihoods for specified outcomes to the question concerned with what decision the firemen would reach

OUTCOME	CONDITION I			CONDITION II	CONDITION III	
	FORESIGHT	HINDSIGHT 6 WEEKS	HINDSIGHT 3 MONTHS		(i) SUMMARY	(ii) DETAILS
Firemen get 30% wage rise	10	5	10	0	2	5
Firemen get 10% + reduced hours	30	25	40	30	30	10***
Firemen get 10% + reduced hours + productivity deal	50	40	40	50	40	20**
Firemen get 10% + guarantee of increases in Nov 1978 & 1979	40	50+	40	50*	50	55
Compromise somewhere between 10% - 20%	40	30	30	40	40	20*

This is the outcome which occurred.

Values with * denote that comparison with appropriate Foresight (Condition I) is significantly different

+ $p < 0.1$; *$p < 0.05$; **$p < 0.025$; ***$p < 0.01$.

Table 9.2 Median Likelihoods for specified outcomes to the question concerned with how long the strike would go on.

OUTCOME	CONDITION I			CONDITION II	CONDITION III	
	FORESIGHT	HINDSIGHT 6 WEEKS	HINDSIGHT 3 MONTHS		(i) SUMMARY	(ii) DETAILS
More than two weeks (but less than 4 weeks)	40	50	40	50	50	30
From 1 - 2 weeks longer	30	25	20	10*	25	20
More than 4 weeks	15	50***	50***	60*	40**	40***
Less than a week	10	10	10	0**	5*	5*

[ ] This is the outcome which occurred

Values with * denote the comparison with appropriate foresight condition is significantly different.

*p < 0.05; **p < 0.025; ***p < 0.01.

(c) Judgements about the Part-time Fire fighters

The analysis next examined judgements about what the Part-time Firefighters would do. From Table 9.3, for the outcome which occurred (carry on as they had been doing), it can be seen than no hindsight assessments were greater than foresight ones. Five of the six were in the opposite direction, however, none of these was significant. Table 9.3 also shows that for outcomes which did not occur, only three of the possible ten foresight/hindsight comparisons were in the predicted direction. Once again neither these or those in the opposite direction were significantly different.

(d) Judgements about the Fire Officers

Table 9.4 reveals that for the outcome which occurred (Fire Officers carry on working and assisting troops and pursue their own pay claim by negotiation) none of the five hindsight assessments were greater than the foresight one. Those in the opposite direction were not significantly different. Table 9.4 also shows that for the outcomes which did not occur, only one of the fifteen possible foresight/hindsight comparison was in the predicted direction. Neither this or those in the opposite direction were found to be significantly different.

(2) Data from Treatments when Outcomes were Generated by Subjects

The analysis next examined foresight and hindsight judgements in the "generate" treatment. The outcomes relating to the decision the F.B.U. would reach were collated from the "generate" treatment in Conditions I, II and III. The collating was done by the author

Table 9.3 Median Likelihood assessments for specified outcomes to the question concerned with the part-time Fire Fighters

OUTCOME	CONDITION I			CONDITION II	CONDITION III	
	FORESIGHT	HINDSIGHT 6 WEEKS	HINDSIGHT 3 MONTHS		(i) SUMMARY	(ii) DETAILS
Come out in support of the Firemen's claim	10	20	20	20	20	20
Carry on as they had been doing	80	80	75	60	60	60
Stop helping troops because of F.B.U. pressure	30	20	20	35	20	30

[ ] This is the outcome which occurred.

Table 9.4 Median Likelihood assessments for specified outcomes to the question concerned with the Fire Officers

OUTCOME	CONDITION I			CONDITION II	CONDITION III	
	FORESIGHT	HINDSIGHT 6 WEEKS	HINDSIGHT 3 MONTHS		(i) SUMMARY	(ii) DETAILS
Come out on strike for their own 40% wage claim	10	20	15	10	10	10
Carry on working and assisting troops and pursue their own claim by negotiation	70	60	65	65	60	70
Come out on strike in support of the Firemen's (FBU) pay claim	5	20	15	15	10	10
Carry on as they are and wait and see what Firemen get before pursuing their own claim	50	50	55	60	60	40

[ ] This is the outcome which occurred

assigning the outcomes generated to one of five categories (the four outcomes in Table 9.5 and a not applicable category). Two other judges independently collated the same outcomes. Correlation, using Kappa (Cohen, 1960) of 0.89 and 0.88 with the author's classifications were obtained. Where there was disagreement the category 2 out of 3 judges had given for the outcome was used. Appendix 9.C summarises outcomes generated by subjects and shows how the author and two judges classified them.

Table 9.5 summarises the median likelihood assessments together with the number of judges who generated each of the outcomes. The number of outcomes each judge generated varied, in Condition I the average was 5.2, in Condition II, 5.3 and in Condition III, 4.8. For the outcome that occurred, four out of the five assessments were greater in hindsight, but not significantly so. Table 9.5 also reveals that for the outcomes which did not occur two hindsight assessments in Conditions I and II were significantly lower in foresight (Compromise somewhere between 10%-20%; Condition II, $U = 19$, $P < 0.05$; Condition III (ii) $U = 5$, $P < 0.025$). No significant differences were found in foresight/hindsight comparisons in Condition I.

(3) Type of Information - "summary"/"details"

Providing different types of information ("summary" and "details") yielded a consistent pattern of results in Condition III (i) and (ii) for both treatments. The medians for the various outcomes are given in Tables 9.1 and 9.5. The outcome that occurred was seen as more likely after judges had read the "details" than the "summary". The difference was significant for the generate treatment ($P < 0.025$;

Table 9.5 Median Likelihood assessments assigned to generated outcomes concerned with the decision the F.B.U. would reach

OUTCOME	CONDITION I			CONDITION II	CONDITION III	
	FORESIGHT	HINDSIGHT 6 WEEKS	HINDSIGHT 3 MONTHS		(i) SUMMARY	(ii) DETAILS
Fireman get 30% wage rise	5 N = 10	10 N = 10	10 N = 10	10 N = 5	20 N = 8	0.05 N = 8
Firemen get 10% + reduced hours	17.5 N = 10	20 N = 10	20 N = 10	30 N = 9	30 N = 6	40 N = 6
Firemen get 10% + guarantee of increases in Nov 1978 & 1979	50 N = 10	60 N = 10	60 N = 10	60 N = 14	30 N = 9	60 N = 9
Compromise somewhere between 10%-20%	50 N = 11	60 N = 11	65 N = 11	10* N = 7	37.5 N = 4	20** N = 4

This is the outcome which occurred.

Values with * denote that comparison with the appropriate before condition is significantly different.

*p < 0.05; **p < 0.025.

T = 2.5, N = 1) (Table 9.5), but not for the specified treatment (Table 9.1). For the outcomes which did not occur, five out of the seven medians were lower after "details" then "summary". Three of these comparisons being significant - "specified": Firemen get 10% and reduced hours, T = 9, N = 11, P < 0.05; Firemen get 10% and reduced hours and productivity deal, T = 0, N = 9, P < 0.05 - and one a trend - "generate": compromise between 10%-20%, T = 0, N = 7, P < 0.1. The results from the generate treatment here need to be treated with some caution because of the low power of the statistical tests due to small sample size.

(4) Attitudes and Judgements

From the "specified" treatment it was possible to categorise certain people in two ways to give the most extreme difference in attitudes

(i) Those who indicated strong support for the Government Pay Policy (6 or greater on the appropriate seven-point scale) and strongly opposed the F.B.U.'s 30% wage claim (2 or less on the appropriate scale).

(ii) Those of strongly opposed the Government's Pay Policy (2 or less) and strongly supported the F.B.U.'s 30% wage claim. (6 or more).

Insufficient numbers of people fell into these categories in the Before conditions consequently they were dropped from the analysis. However, in the hindsight conditions (I, after 6 weeks; II and III(i)) there were sufficient numbers for statistical analysis to be performed.

Table 9.6 Summary of median Likelihood assessments for specified outcomes for subjects indicating strong support for the Government Pay Policy and strong support for the F.B.U. wage claim

QUESTION AND OUTCOMES	SUPPORT GOVT	SUPPORT F.B.U.
DECISION F.B.U. WOULD REACH	N=13	N=12
1. Firemen get 10% and guarantee of rises in November 1978 and 1979	60	50
2. Firemen get 30% wage rise	0	3.5
3. Firemen get 10% and reduced hours	30	25
4. Firemen get 10% and reduced hours and productivity	40	32.5
5. Compromise somewhere between 10%-20%	17.5	35
HOW LONG STRIKE WOULD GO ON FOR		
1. More than 4 weeks (from week 4)	40	50
2. More than 2 weeks (but less than 4)	45	50
3. From 1-2 weeks longer	40	20
4. Less than a week	10	2.5
PART-TIME FIRE FIGHTER		
1. Carry on as they had been doing	40	69*
2. Come out on strike in support of the Firemen's claim	30	10
3. Stop helping troops because of F.B.U. pressure	50	20
FIRE OFFICERS		
1. Carry on working and assisting troops and pursue their own claim by negotiation	75	55*
2. Come out on strike for their own 40% wage claim	10	15
3. Come out on strike in support of the F.B.U. wage claim	10	10
4. Carry on, wait and see what F.B.U. get before pursuing own wage claim	50	62.5

For each question Outcome 1 is the one which happened.

*$P < 0.05$

Table 9.6 summarises the median likelihood assessments for the four questions and their outcomes for subjects in the above two categories. The column labelled "support Govt" refers to category (i) above, that labelled "support F.B.U." to category (ii) above. For the question concerned with the decision the F.B.U. would reach the two most favourable outcomes for the F.B.U. (30% wage rise, and compromise between 10-20%) were perceived as more likely by those in category (i) than those in category (ii). These differences were not significant however. Those outcomes regarded as favouring the Government (10% and reduced hours, 10% and reduced hours and productivity deal) were seen as more likely by those in category (i) and (ii). Again, there were not significant differences.

No significant differences were found for the outcomes concerned with the Question concerning how long the strike would continue from Week 4. The outcome which occurred for the questions concerned with the part-time Fire Fighters and Fire Officers was seen as more likely in category (ii) in the case of the former ($U = 37$, $P < 0.05$, two-tailed) and more likely in category (i) for the latter question ($U = 43$, $P < 0.05$, two-tailed). For the outcomes which did not occur for these two questions no significant differences were found.

9.4. DISCUSSION

The firstissue this experiment was designed to explore was whether the bias reported by Fischhoff and his colleagues is present in judgements about contemporary socio-political events. Results offer only partial support for the claim that possession of outcome knowledge leads to an inappropriate "knew-it-all-along" attitude. The support

comes from the findings in the "specified" treatment for the questions, what decision the F.B.U. would reach and how long the strike would continue. The two outcomes which occurred were perceived as more likely in hindsight than foresight; the outcomes which did not occur were seen as less likely in hindsight. No hindsight bias was detected, however, with "specified" outcomes concerning the remaining two issues, namely the part-time Fire Fighters and the Fire Officers. Subjects who made both foresight and hindsight assessments showed much less of a "knew-it-all-along" attitude confirming the findings of Fischhoff and Beyth (1975) and Fischhoff (1977). This is probably due, as they argue, to the logical role of memory, in that some subjects will explicitly remember their own probability assessments.

The experiment failed to replicate Fischhoff and Beyth's (1975) findings that hindsight bias became more pronounced the longer the period of time which elapsed between the outcome of the event being known and hindsight judgements being made. No significant differences were found between Hindsight, 6 weeks and 3 months in condition I. Furthermore, only two of the five possible comparisons was in the direction that would be predicted from Fischhoff and Beyth's research (see Table 9.1).

The second issue was whether the bias would be influenced by the way in which probability assessments were elicited. People asked to "generate" their own outcomes showed little or no bias in any of the three conditions. In Condition I this may be partly attributable to the discrepancy in number of outcomes assigned probability assessments in foresight. In the "specified" treatment subjects gave assessments

to outcomes concerned with four aspects of the strike, making seventeen in all. In the "generate" treatment the average number of outcomes produced was just over five. Thus, memory may account for the lack of foresight/hindsight differences in the latter treatment. However, this problem does not arise in Conditions II and III as subjects here made only hindsight judgements. These results then, provide some support for the hypothesis that the "generate" strategy may aid judges accurately reconstructing a foresight state. The results cannot be taken as providing strong support for this hypothesis because of the reduced power of the statistical tests due to small sample size.

The third issue was whether the bias would be influenced by the type of information. The impact of "details" (see p. 225) in Condition III(ii) was to make the outcomes which had occurred seem more likely than after reading the "summary" (Condition III(i)). Conversely, outcomes which did not occur were perceived as less likely after "details" than "summary" information. Both effects were the opposite to that predicted.

An explanation for this comes from recent research on biases and errors in attribution (Ross, et al, 1975; Ross, 1977; Ross et al 1977) which was discussed in detail, in Chapter 2. Briefly to repeat these findings Ross et al (1977) found that giving judges causal explanations as to why a social event turned out as it did increased perceptions of likelihood in that direction. They concluded that "identifying antecedents to explain an event increases the event's subjective likelihood" (p. 826).

In the context of the experiment reported here it can be seen that providing the "details" is equivalent to identifying antecedents. The subject's cognitions of the Firemen's strike are biased because the "details" allow judges to construct causal explanations of why what happened did, and why what did not happen did not. Consequently, the outcomes which happened were perceived as more likely, and those which did not were perceived as less likely.

The fourth, and final, issue this experiment was designed to explore concerned the effects of subjects' attitudes upon subjective likelihood assessments. It was predicted that those who strongly supported the Government Pay Policy and strongly opposed the F.B.U. wage claim would perceive outcomes favouring the Government was more likely and those favouring the F.B.U. as less likely than judges who strongly held the opposite views. Conversely, those favouring the F.B.U. and opposing the Government would perceive the same outcomes in the opposite way. No substantial support was found for this hypothesis. Inspection of the medians in Table 9.6 did reveal the likelihood assessments for the outcomes associated with the decision the F.B.U. would reach to be in this direction. However, as no significant differences were found here little confidence can be placed in the reliability of them. This experiment was not designed with attitudes as a major theme and it may have been because strong supporters of both sides were not actively sought as subjects that the above non-significant findings occurred.

Experiment 7, reported in Chapter 11 of this thesis, rectifies. such shortcomings by investigating judgements in foresight and hindsight

made by Labour and Conservative Party members to various outcomes of the British General Election of 1979.

In summary, the experiment reported here has provided some support for hindsight bias and indications of the conditions under which it might not occur. The research to date, however, has not demonstrated that such a bias occurs when people are asked to make judgements about aspects of their own personal lives. The following experiment was designed with this in mind.

CHAPTER 10

EXPERIMENT 6[1]: ESTIMATING THE OUTCOME OF A PREGNANCY TEST: WOMEN'S JUDGEMENTS IN FORESIGHT AND HINDSIGHT

10.1 INTRODUCTION

The study to be reported here investigates the extent to which hindsight bias will be found when people are asked to make subjective likelihood assessments which require them to look at certain "historical" aspects of their own lives. The area chosen for study was the uncertain event of pregnancy.

Most of the research conducted to date has required likelihood assessments to be made to outcomes concerning either (1) event descriptions "sufficiently familiar to permit intelligent responses and sufficiently unfamiliar to rule out the possibility of subjects knowing what really happened" (Fischhoff, 1975a, p. 290) or (2) contemporary socio-political events. The main drawback of these approaches is that the material may have little or no relevance for the judges; i.e. making judgements and predictions related to salient features of their personal lives. Experimental studies will almost certainly be required to explain the nature of hindsight bias, but if we are to be confident that there is a "real life" effect which requires explanation, it is important to demonstrate hindsight bias

---

1. The data for this experiment was collected by Kay McKenna for her third year project. Derek Rutter supervised the project with some assistance from the author. The author re-analysed the data and wrote the first draft of the paper which is to appear in the British Journal of Social and Clinical Psychology. The first draft was commented upon, and modified by, Derek Rutter and Ian Morley.

in non-laboratory settings. To this end asking women to make foresight and hindsight assessments to the outcome of a pregnancy test will provide invaluable information about the extent of hindsight bias in everyday life.

Two procedures were used to investigate this (1) Women who suspected they might be pregnant were asked to predict, on first contact with a branch of the British Pregnancy Advisory Service, the likelihood (foresight) that they were infact pregnant. A second (hindsight) assessment, asking women to remember or reconstruct their original assessment, was obtained once the women had been given the result of the pregnancy test. (2) Women who know the results of their tests were asked to remember or reconstruct assessments they would have made some time previously (hindsight).

It was predicted that those who knew the result of the pregnancy test (hindsight) would perceive that outcome as more likely. That is, those who received a positive result should see their chances of being pregnant as more likely in hindsight than foresight, conversely those receiving a negative result should, in hindsight, see their chance of being pregnant as less likely than in foresight.

10.2 METHOD

PROCEDURE

Women normally make an initial contact with pregnancy advisory services by telephone or in person. This study included two "telephone" conditions and one "in person" condition, and was conducted at the Coventry branch of the British Pregnancy Advisory Service (B.P.A.S.).

Conditions 1 and 2 included those subjects whose initial contact was a telephone call to the clinic to ask about the possibility of having a pregnancy test. Where this was possible an appointment was made for the test to be carried out at the clinic at a later date. Condition 3 consisted of women who made contact by calling at the bureau to request a test.

Condition 1. Half the telephone callers were randomly assigned to this condition. During the initial telephone conversation, the subject was asked a number of routine questions for the clinic's records. She was then asked to predict, using a scale from 0 to 100 ("where 0 means you are definitely not pregnant, 100 means you definitely are, and don't know is somewhere around the middle), how likely she thought she was to be pregnant. Approximately 4 to 5 days after this initial contact, the subject attended the clinic to have her test and she was told the outcome a few minutes later. After receiving the result, she was asked to remember or reconstruct the prediction she had made originally using the same scale.

Condition 2. Half the telephone callers were randomly assigned to this condition. During the initial telephone conversation, a pregnancy test appointment was made. Only information pertinent to the test was requested, and the subject was not asked how likely she thought she was to be pregnant. The second contact was made in person at the clinic, at which point the subject was given her pregnancy test and was then told the results a few minutes later. After she had heard the result, she was asked to remember or reconstruct how likely she thought she was to be pregnant at the time of the initial

contact 4 to 5 days earlier.

Condition 3. This condition consisted of those women who came to the clinic in person to request a pregnancy test. No prior appointment had been made, and the subject was simply given the test and told the outcome a few minutes later. After she had heard the result, she was asked to remember or reconstruct how likely she had thought she was to be pregnant when she first decided to have the test.[2]

SUBJECTS

Altogether, 60 women were included, 20 in each condition. Ten in each condition proved to have a positive result to the pregnancy test and ten had a negative result. Certain criteria for inclusion had to be met, and these were intended to ensure that subjects seriously thought they might be pregnant.

(a) It was at least 41 days from a last period at the time of the test appointment. Those women who simply required information about pregnancy testing but did not wish to have a test were excluded.

(b) The woman was prepared to attend the clinic to have her urine sampled.

(c) She had not been sterilised.

(d) She had had no previous pregnancy test in the last 41 days.

Note on the Pregnancy Test

The pregnancy test was performed by means of an immunological slide analysis of the urine sample (Kerber et al, 1970). It was

---

[2] About half the women who contacted the agency without prior appointment were included as subjects in Condition 3. This was done to balance the 50% of women who failed to turn up to a previously made appointment (as in Conditions 1 and 2).

conducted after 41 days or more had elapsed since the last menstrual period because the technique is unreliable within 41 days (Tietz, 1965; Mayo and Thompson, 1965). For a more detailed discussion see Cabrera (1969).

Post-experimental follow up

Many subjects whose initial contact was made by telephone failed to keep their pregnancy test appointment. To test whether applicants might be dropping out because of the questionnaire - so leading to an unrepresentative sample - control condition was run in which the questionnaire was excluded. Of those who were asked the question, 51.3% returned for the pregnancy test, against 46% of those who were not asked the questions. Indications are, therefore, that the experimental questionnaire did not discourage women from returning, and that the experimental sample is unlikely to have been unrepresentative of the clinic's population.

Questionnaire and instructions

The routine questions asked of all women for the clinic's records were as follows:

Name
Age
Date of contact
Status: Married Single Other
L.M.P. Date
Approximate size in weeks
Number of previous pregnancies
Contraceptive used
Do you hope you are pregnant?
Are you trying to become pregnant?

Subjects allocated to Condition 1 before were asked to make subjective likelihood assessments in accordance with the following

instructions:

Personal contact after

Name
Test result: Positive Negative

I would like you to remember back to when we spoke last. Can you remember what you thought the result of the test would be?

Imagine, say, you've got a scale from 0-100, where 0 means you definitely are not pregnant, 100 means you definitely are, and don't know is somewhere around the middle. What score would you have put on your chances of being pregnant?

Subjects in Condition II were provided with similar instructions. Women in Condition III, because they had not made prior contact with the clinic by telephone, were provided with slightly different instructions. These ran as follows:

I would like you to remember back to when you first decided to have a pregnancy test

(1) When did you decide?
(2) Can you remember what you thought the result of the test would be?

Imagine, say, you've got a scale from 0-100 where 0 means you definitely are not pregnant, 100 means you definitely are, and don't know is somewhere around the middle. What score would you have put on your chances of being pregnant?

10.3 RESULTS

To test the prediction, that those who knew the result of the pregnancy test (hindsight) would perceive that outcome as more likely than those asked to make subjective likelihood assessments before they knew the result (foresight), two sets of comparisons between conditions were made. First, for condition 1, in which each subject gave both a foresight and hindsight assessment, the two assessments

were compared by the Wilcoxon Matched pairs signed-ranks test. Before and After assessments showed no significant differences both for women who had a positive result to the pregnancy test (Wilcoxon T = 0, N = 3, n.s.) and for those who had a negative result (Wilcoxon t = 9, N = 4, n.s.).

In the second analysis, the reconstructed assessments in Condition 2 and the reconstructed assessments in condition 3 were compared with the Before assessments in condition 1 using the Mann Whitney U test. Women who received a positive result to the pregnancy test in both Condition 2 and 3 perceived the outcome as more likely than women in foresight (Mann Whitney U = 27, $P < 0.05$; Mann Whitney U = 21, $P < 0.025$, respectively). Women who received a negative result in conditions 2 and 3 showed no difference in likelihood assessments than women in foresight (Mann Whitney U = 44.5, n.s.; Mann Whitney U = 48, n.s., respectively).

The results summarised in Table 10.1 also indicated that, whether in foresight or hindsight, women whose test proved positive saw themselves as more likely to be pregnant than those whose result was negative. To test for this comparisons were made between assessments given by women who had a positive and negative result in each condition using the Mann Whitney U test. All four comparisons proved to be significant (1 Before (+) - 1 Before (-): U = 24, $P < 0.05$; 1 After (+) - 1 After (-): U = 19.5, $P < 0.025$; 2 After (+) - 2 After (-): U = 24.5, $P < 0.05$; 3 After (+) - 3 After (-): U = 13, $P < 0.01$). The median assessments for these conditions are depicted graphically in Figure 10.1.

Figure 10.1 Median likelihood assessments (%) for each condition for positive and negative pregnancy test results.

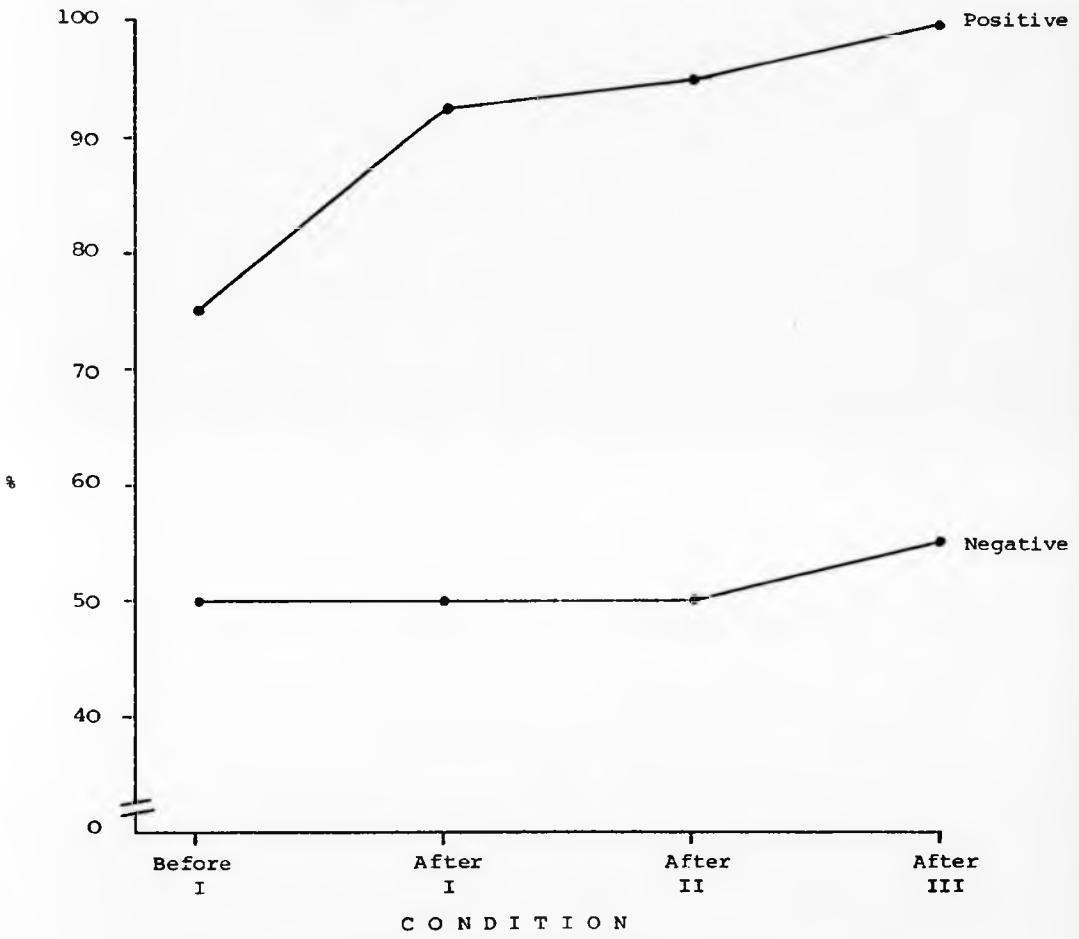


Table 10.1: Summary of Median Likelihood Assessments and Biographical Details

CONDITION	RESULT OF PREGNANCY TEST	N	MEDIAN LIKELIHOOD ASSESSMENT	DAYS BETWEEN CONTACT AND TEST	MEAN AGE	PREVIOUS PREGNANCY		CONTRACEPTION	
						YES	NO	YES	NO
1 BEFORE	POSITIVE	10	75	5.5	23.6	5	5	5	5
1 AFTER	POSITIVE	10	92.5						
1 BEFORE	NEGATIVE	10	50	4.7	33.4	8	2	8	2
1 AFTER	NEGATIVE	10	50						
2 AFTER	POSITIVE	10	95	4.1	25.1	6	4	5	5
2 AFTER	NEGATIVE	10	50	4.5	30.2	7	3	6	4
3 AFTER	POSITIVE	10	100	-	25.7	5	2	2	8
3 AFTER	NEGATIVE	10	55		28.4	5	5	5	5

In summary, then, there were two main findings. First, for women whose test proved positive, there was evidence that those who knew the result of the pregnancy test (hindsight) perceived that outcome as more likely than those asked to make the assessment before they knew the result (foresight), though this was only so when different women made hindsight and foresight judgements, and not when repeated measures were taken on the same women. Women whose result was negative showed no effect and were undecided about the outcome whenever they made the assessment. Second, consistently throughout the study, women whose test proved positive made higher assessments than those whose tests proved negative. Typically, those who were not pregnant were undecided (with a median assessment of 50%) while those who were pregnant made much higher assessments (with a median of 92.5). These findings suggest that who proved to be pregnant have some insight.

In order to understand how the pregnant group may have obtained this insight, a number of comparisons between the positive and negative groups were made using the personal information that had been collected about the women. This information is summarised in Table 10.1. The particular variables examined were (a) whether the woman had been pregnant before, (b) her age, (c) whether contraceptives were being used. Each variable was examined by chi-squared (Siegel, 1956, p.107), combining data for all three conditions.

Neither previous pregnancy (χ^2 = 0.625) nor age proved significant. However, for contraceptives/no contraceptives, a trend which approached significance was revealed (χ^2 = 2.4, df = 1,). That is, there was some indication that those who were not using contraceptives thought they were more likely to be pregnant than those who were.

A further analysis was then performed, comparing on the one hand those who were using contraceptives with on the other those who were not, combined with women who reported contraceptive failure. The "contraceptive failure" group included those women who had specifically reported that their method of contraception had failed; for example, they had forgotten to take their pill or there had been some contraceptive failure. This revealed a trend at the 10% level (χ^2 = 3.5, df = 1).

Taken together, these background data account perhaps for some of the difference between the positive and negative groups, although much of the variance remains unexplained.

In order to understand how the pregnant group may have obtained this insight, a number of comparisons between the positive and negative groups were made using the personal information that had been collected about the women. This information is summarised in Table 10.1. The particular variables examined were (a) whether the woman had been pregnant before, (b) her age, (c) whether contraceptives were being used. Each variable was examined by chi-squared (Siegel, 1956, p.107), combining data for all three conditions.

Neither previous pregnancy (χ^2 = 0.625) nor age proved significant. However, for contraceptives/no contraceptives, a trend which approached significance was revealed (χ^2 = 2.4, df = 1,). That is, there was some indication that those who were not using contraceptives thought they were more likely to be pregnant than those who were.

A further analysis was then performed, comparing on the one hand those who were using contraceptives with on the other those who were not, combined with women who reported contraceptive failure. The "contraceptive failure" group included those women who had specifically reported that their method of contraception had failed; for example, they had forgotten to take their pill or there had been some contraceptive failure. This revealed a trend at the 10% level (χ^2 = 3.5, df = 1).

Taken together, these background data account perhaps for some of the difference between the positive and negative groups, although much of the variance remains unexplained.

10.4 DISCUSSION

The purpose of this experiment was to explore, in a "real-life" setting, Fischhoff's suggestion that when outcomes of events are seen in hindsight, they are perceived as more likely than when those same events are seen in foresight. The uncertain event chosen for examination was the outcome of a pregnancy test, and only partial support was found for hindsight bias.

The present findings suggest that hindsight bias is a less pervasive phenomena outside the laboratory in "real life". This is so because the bias was present when a positive outcome to the pregnancy test was obtained but not when the test proved negative. Janis and Mann (1977) have argued that, when one is investigating decision-making and judgement under uncertainty, it is essential to construct experiments which have relevance and consequence for subjects. Much previous research on judgement under uncertainty does not have this feature. The present study, although not having consequence, certainly does have obvious relevance. Subjects are being asked to give a formal assessment to their personal feelings and cognitions about something of the greatest importance - the likelihood of them being pregnant.

The above considerations are important because they provide a clear framework by which the findings of this experiment can be explained. The fact that contact was made with a pregnancy advisory service implies a strong suspicion of pregnancy. Those women who received a positive result had their suspicions confirmed and were biased to seeing this outcome as more likely in hindsight. In

contrast, for those who received a negative result, attendance at the agency served to counteract the effect of outcome knowledge, and so they showed no bias.

In conclusion, this experiment has demonstrated that before we can generalise findings from the laboratory, or where the stimulus material does not have personal relevance to subjects, to everyday life research is need in applied settings to determine what situational variables, for example, may mediate the effect. In this experiment, hindsight bias has been shown to occur only when a person's suspicions in foresight have been later confirmed. Where grounds for the suspicions have been removed hindsight bias did not occur. The experiment has shown this to be the case when people seek an answer from some objective source or authority, further research is required to discover whether this principle holds regardless of the objectivity of the "authority". Finally, the experiment indicates that an experiment designed to have relevance for subjects may yield important indications as to what conditions encourage or reduce hindsight bias, the next experiment to be reported in this thesis was designed with this in mind.

CHAPTER 11 .

EXPERIMENT 7: THE BRITISH GENERAL ELECTION OF MAY 1979: JUDGEMENTS IN FORESIGHT AND HINDSIGHT BY LABOUR AND CONSERVATIVE PARTY MEMBERS

11.1 INTRODUCTION

The British General Election of the 3rd May, 1979 came about as the result of a successful no confidence motion tabled by the Conservatives. The previous General Election in October 1974 gave Labour an overall majority of only three seats. This soon dwindled to a position where they had no overall majority. The no confidence motion in April, 1979 was successful because the minority parties voted with the Conservatives.

In the run up to the General Election of May, 1979 the opinion polls consistently predicted an overall majority to Conservative. This proved to be the case: the result gave the Conservatives an overall majority of 42 seats. Labour lost nearly this many, and over the whole country the conservatives achieved a swing of over 5 per cent away from Labour. The Liberals lost three seats and the Scottish Nationalist Party, possibly providing the biggest surprise of the election, were decimated. They retained only two seats, losing nine. The victory by the Conservative Party gave Britain its first ever woman prime minister.

Using Labour and Conservative party members as subjects, the experiment to be reported here made use of the General Election to investigate (1) hindsight bias, and (2) how attitudes and values

affect subjective likelihood assessments.

With respect to the first issue it was hypothesised that those who knew the result of the election would see the outcomes which occurred as more likely, and those which did not occur as less likely, than subjects in foresight. With respect to the second issue, it was hypothesised that outcomes favourable to a particular party would be seen as more likely, in both foresight and hindsight, by that party than the other one. This prediction in conjunction with the former one should also mean, as a consequence, that hindsight bias should be greater, for the outcomes which occurred, for party members where the result is in their favour. Conversely, the party who suffered defeat should exhibit less bias for the outcomes which occurred.

The use of active Labour and Conservative Party members as subjects in this experiment extends the theme of the experiment reported in the previous chapter. There it was found that designing experiments with material relevant to subjects provided important information concerning the conditions which may encourage or reduce hindsight bias. The present experiment investigates how attitudes, wishes, commitment or values (i.e. Party affiliations) may interact with outcome knowledge to affect subjective likelihood assessments.

McGregor (1938), Cantril (1938), as discussed in Chapter 9, both showed that people have a tendency to predict the events they would like or wish to see come about. More recently Slovic (1966) and Sevon (1975) have shown that the value a person attaches to an event affects likelihood assessments. Sevon (1975), for example, showed

that people who highly valued certain polical outcomes perceived them as more likely than people who held no opinion or felt neutrally towards the same outcomes.

The aim of the present study was to extend these findings to look at how hindsight judgements may be affected by what people would have liked to have happened. The second hypothesis, then, was derived from the above findings.

11.2 METHOD

SUBJECTS

40 subjects, 20 active members of the Whitnash Conservative Party and 20 active members of the Coventry Labour Party, took part in the experiment.

Conservative subjects were obtained by approaching the chairman of the Whitnash Conservative Party. After briefly describing the project to him he subsequently asked his members if they would fill in a questionnaire. Those who agreed were given the questionnaire to complete at the appropriate time.

Labour subjects were obtained by approaching a committee member of the Coventry Branch of the Labour Party. Again a short outline of the project was related to him. He then asked other party members if they would take part in the study. The general view of those who agreed was that they would only do so if anonymity was ensured. With this restriction Labour subjects were given the questionnaire to complete at the appropriate time.

PROCEDURE

20 subjects, 10 Labour and 10 Conservative were assigned to a foresight condition and 20 subjects, 10 Labour and 10 Conservative, to a hindsight condition. In both conditions subjects assessed the likelihood of outcomes to seven aspects of the General Election:

1. The result of the general election (a) likelihood assessments
   (b) estimates of surprise
2. The turnout
3. The swing to Labour or Conservative
4. The number of seats majority to Labour or Conservative
5. The Liberals (a) % of votes polled
   (b) swing from/to Liberal
   (c) number of seats
6. The National Front
7. The Scottish Nationalist Party.

In the foresight condition both Labour and Conserative subjects filled out a 10-page questionnaire during the second week before the General Election (i.e. between 23rd - 30th April). Before assigning likelihood assessments to the outcomes related to the seven questions subjects read through an information sheet. This provided details about the result of the General Election of October, 1974 and information about the forthcoming General Election. This information was derived from Butler and Kavanagh (1975) and Mckie, Cook and Phillips (1978), it is reproduced in Appendix 11A.

In the hindsight condition subjects filled out a similar 10-page questionnaire during the 5th and 6th week after the day of the General Election (i.e. between 6th and 16th June). Before filling out the questionnaire subjects were provided with an information sheet summarising the results of the General Election. This was derived from the "Guardian" newspaper of 5th May, 1979 and is reproduced in Appendix 11B.

INSTRUCTIONS

Labour and Conservative subjects in the foresight condition received the following instructions on the cover sheet of the questionnaire:

> In this questionnaire we are interested in your predictions concerning various aspects of the forthcoming General Election. Would you please read through the information relating to the last General Election and information surrounding the forthcoming one. Having done this would you then answer the questionnaire. Thank you.

Subjects in hindsight received the following instructions on the cover sheet of their questionnaire:

> In this questionnaire we are interested in how you thought various aspects of the recent General Election would turn out. Although the results are known it is still of interest to know how people thought it would turn out. This being the case, I am asking you to cast your mind back to about a week before voting day (May 3rd) and try to assess how likely you thought each of the various outcomes listed in the questionnaire seemed to you before the results were known. That is, try to think back to before May 3rd and answer the questionnaire as you believed you would have done then.
>
> You will find in what follows an information sheet stating the results of the General Election. Furthermore, with each question you will see an * before one of the possible outcomes, this indicates that it is the one which happened. This information has been provided as people will vary in how much detail they know about the General Election result. This being so I have decided to provide all relevant information so that all people who answer this questionnaire know the same amount.
>
> Would you please read through the information sheet then procede with answering the questionnaire.

QUESTIONNAIRE

Subjects in both conditions were asked to make subjective likelihood assessments to seven aspects of the General Election, these were given earlier. The number of outcomes specified to each question ranged from four to nine. For example, one question (Number 3) asked subjects what they think/thought the swing at the General Election will/would be.

They had to indicate, on appropriate scales, how likely they thought the following outcomes were:

1. Swing to Labour of more than 10%
2. Swing to Labour of between 5-10%
3. Swing to Labour of between 2-5%
4. Swing to Labour of between 0.1-2%
5. No swing to Labour or Conservative
6. Swing to Conservative of between 0.1-2%
7. Swing to Conservative of between 2-5%
8. Swing to Conservative of between 5-10%
9. Swing to Conservative of more than 10%

Specifying so many outcomes to some questions was necessary for two reasons: (1) to ensure that the one that occurred would be present; (2) to provide a reasonable differentiation of the possibilities. Having so many outcomes meant that it was unreasonable and unrealistic to ask subjects to provide subjective likelihood assessments in percentages and ensure that they totalled 100% for any one question. Consequently, subjects were asked to make assessments in one of three ways, depending on the question. (1) in percentage form, the total adding to 100 per cent; this was only for Question 1 where only five outcomes were specified. (2) in terms of degree of surprise, expressed on a seven-point scale: this was for the same outcomes as in Question 1. (3) on a seven-point scale ranging from very likely (7) to not very likely (1); this was used for questions 2-7; subjects were asked to indicate these assessments by placing a vertical line on the scale in a position which best accorded with their subjective likelihood.

After each question and before the specified outcomes, a summary of relevant information, a short description of what the question was getting at, and an explanation of any technical terms was given.

This was provided to ensure that all subjects fully understood the question and associated outcomes. Appendix 11.C reproduces the questionnaire given to subjects in hindsight. The only differences between this and the one given in foresight was that the wording of the questions was changed from past to future tense and, obviously, subjects were not informed of the result as it was not known at the time.

MEASUREMENT

A scale value for each outcome for each question had to be determined where subjects made likelihood assessments on seven-point scales. When designing the questionnaire the length of each scale was ensured to be the same (48 m.m.). Conversion to a scale value was achieved by first measuring the distance of the subject's response from the "not very likely" end. Having already determined the length of each scale point the following formula was applied to determine the scale value (SV):

$$SV = 1 + \frac{n}{m}$$

where n = response by subject (mm)

m = length of scale point (mm)

11.3 RESULTS

In order to discover how outcome knowledge and party affiliations influenced subjective likelihood assessments two sets of analyses, using the Mann Whitney U test, were carried out. First comparisons between Labour subjects in foresight and hindsight, and Conservative subjects in foresight and hindsight. Second, comparisons between Labour and Conservative in foresight and Labour and Conservative in

hindsight. One-tailed tests were used when differences were in the predicted direction, two-tailed when in the opposite direction.

(a) Foresight/Hindsight

Tables 11.1 to 11.10 summarise the median likelihood assessments given to the outcomes associated with each of the questions about the General Election. The Tables also show the Mann Whitney U statistic for foresight/hindsight comparisons and whether the differences are reliable or not. In what follows this section provides a short question by question description of the results, followed by an overall summary of them.

Q.1. The result of the General Election

(a) Likelihood assessments. Inspection of Table 11.1 reveals that, for the outcome which occurred (overall majority to Conservative), there was no significant difference between assessments for either Labour or Conservative subjects. None of the medians, for the outcomes which did not occur, were in the predicted direction. The one significant result (majority to Conservative: Conservative subjects) was in the opposite direction to that predicted by hindsight bias.

(b) Estimates of surprise: Table 11.2 shows that the foresight/hindsight medians for the outcome which occurred are both in the opposite direction (i.e. more surprised in hindsight), but not significantly so. For the outcomes which did not occur, 5 of the 8 hindsight medians were greater than foresight ones, two significantly so.

Q.2. The turnout

Table 11.3 shows that for the outcome which occurred (between 73-78%) there was no significant difference for Labour subjects.

Table 11.1 Summary of Medians and statistical analysis for likelihood assessments (%) given to Q1 - The result of the General Election

OUTCOME	FORESIGHT-HINDSIGHT					LAB-CON		
	CONDITION	MEDIAN FORE	MEDIAN HIND	U	SIG	CONDITION	U	SIG
Overall Maj. to Lab	Lab	12.5	12.5	48	NS	FORESIGHT	42.5	NS
	Con	7.5	10.0	37	NS	HINDSIGHT	29.0	NS
Overall Maj.to Con	Lab	17.5	25.0	37	NS	FORESIGHT	4.0	P<.01
	Con	52.5	40.0	35	NS	HINDSIGHT	34	NS
Majority to Lab	Lab	32.5	37.5	42	NS	FORESIGHT	17	P<.01
	Con	15.0	17.5	40.5	NS	HINDSIGHT	25.5	P<.05
Majority to Con	Lab	30.0	30.0	42.5	NS	FORESIGHT	31.5	NS
	Con	20.0	34.5	19	P<.01	HINDSIGHT	32.5	NS
Other Outcome	Lab	0	0	50	NS	FORESIGHT		NS
	Con	0	0	50	NS	HINDSIGHT		NS

[ ] This is the outcome which occurred

Table 11.2 Summary of medians and statistical analysis for surprise estimates given to Q1.

OUTCOME	FORESIGHT-HINDSIGHT					LAB-CON		
	CONDITION	MEDIAN FORE	MEDIAN HIND	U	SIG	CONDITION	U	SIG
Overall Maj. to Lab	Lab	3.69	5.38	25	P<.05	FORESIGHT	20	P<.025
	Con	6.25	6.50	41	NS	HINDSIGHT	23	P<.025
Overall Maj.to Con	Lab	3.82	4.88	43	NS	FORESIGHT	11.5	P<.01
	Con	1.75	2.06	28	NS	HINDSIGHT	19.5	P<.025
Maj. to Lab	Lab	2.20	3.25	42	NS	FORESIGHT	17.0	P<.01
	Con	5.82	5.00	49.5	NS	HINDSIGHT	26.5	P<.05
Maj. to Con	Lab	2.20	2.75	39	NS	FORESIGHT	36	NS
	Con	2.81	2.75	49.5	NS	HINDSIGHT	46.5	NS
Other outcome	Lab	6.88	7.00	26.5	P<.05	FORESIGHT	47	N
	Con	6.94	6.75	44	NS	HINDSIGHT	21	P<.05

[ ] This is the outcome which occurred

Table 11.1 Summary of Medians and statistical analysis for likelihood assessments (%) given to Q1 - The result of the General Election

OUTCOME	FORESIGHT-HINDSIGHT					LAB-CON		
	CONDITION	MEDIAN FORE	MEDIAN HIND	U	SIG	CONDITION	U	SIG
Overall Maj. to Lab	Lab	12.5	12.5	48	NS	FORESIGHT	42.5	NS
	Con	7.5	10.0	37	NS	HINDSIGHT	29.0	NS
Overall Maj.to Con	Lab	17.5	25.0	37	NS	FORESIGHT	4.0	P<.01
	Con	52.5	40.0	35	NS	HINDSIGHT	34	NS
Majority to Lab	Lab	32.5	37.5	42	NS	FORESIGHT	17	P<.01
	Con	15.0	17.5	40.5	NS	HINDSIGHT	25.5	P<.05
Majority to Con	Lab	30.0	30.0	42.5	NS	FORESIGHT	31.5	NS
	Con	20.0	34.5	19	P<.01	HINDSIGHT	32.5	NS
Other Outcome	Lab	0	0	50	NS	FORESIGHT		NS
	Con	0	0	50	NS	HINDSIGHT		NS

This is the outcome which occurred

Table 11.2 Summary of medians and statistical analysis for surprise estimates given to Q1.

OUTCOME	FORESIGHT-HINDSIGHT					LAB-CON		
	CONDITION	MEDIAN FORE	MEDIAN HIND	U	SIG	CONDITION	U	SIG
Overall Maj. to Lab	Lab	3.69	5.38	25	P<.05	FORESIGHT	20	P<.025
	Con	6.25	6.50	41	NS	HINDSIGHT	23	P<.025
Overall Maj.to Con	Lab	3.82	4.88	43	NS	FORESIGHT	11.5	P<.01
	Con	1.75	2.06	28	NS	HINDSIGHT	19.5	P<.025
Maj. to Lab	Lab	2.20	3.25	42	NS	FORESIGHT	17.0	P<.01
	Con	5.82	5.00	49.5	NS	HINDSIGHT	26.5	P<.05
Maj. to Con	Lab	2.20	2.75	39	NS	FORESIGHT	36	NS
	Con	2.81	2.75	49.5	NS	HINDSIGHT	46.5	NS
Other outcome	Lab	6.88	7.00	26.5	P<.05	FORESIGHT	47	N
	Con	6.94	6.75	44	NS	HINDSIGHT	21	P<.05

This is the outcome which occurred

Table 11.3 Summary of results for Q.2 - The turnout

OUTCOME	FORESIGHT-HINDSIGHT					LAB-CON		
	CONDITION	MEDIAN FORE	MEDIAN HIND	U	SIG	CONDITION	U	SIG
Greater 78%	Lab	3.81	4.57	49	NS	FORESIGHT	26	P<.05
	Con	1.75	4.00	17.5	P<.01	HINDSIGHT	49	NS
Between 73-78%	Lab	6.63	5.94	35.0	NS	FORESIGHT	19	P<.025
	Con	3.06	5.75	20.0	P<.025	HINDSIGHT	48.5	NS
Between 71-73%	Lab	4.44	4.62	47.0	NS	FORESIGHT	49.5	NS
	Con	4.38	5.00	36.5	NS	HINDSIGHT	38	NS
Between 68-71%	Lab	3.50	3.50	46.0	NS	FORESIGHT	41.5	NS
	Con	3.13	3.00	35.0	NS	HINDSIGHT	48	NS
Less 68%	Lab	1.63	2.31	28.5	NS	FORESIGHT	45	NS
	Con	1.56	1.81	35.5	NS	HINDSIGHT	44	NS

Table 11.4 Summary of results for Q3 - The Swing

OUTCOME	FORESIGHT-HINDSIGHT					LAB-CON		
	CONDITION	MEDIAN FORE	MEDIAN HIND	U	SIG	CONDITION	U	SIG
To Lab > 10%	Lab	1.13	1.68	20	P<.025	FORESIGHT	31.5	NS
	Con	1.25	1.63	34.5	NS	HINDSIGHT	45	NS
To Lab 5-10%	Lab	1.18	2.38	9	P<.01	FORESIGHT	28.5	NS
	Con	1.63	1.63	49.5	NS	HINDSIGHT	17	P<.01
To Lab 2-5%	Lab	2.14	4.06	21.5	P<.025	FORESIGHT	49.5	NS
	Con	2.13	1.87	45	NS	HINDSIGHT	4	P<.01
To Lab 0.1-2%	Lab	4.88	5.19	49.5	NS	FORESIGHT	14	P<.01
	Con	1.87	2.19	47.0	NS	HINDSIGHT	10	P<.01
No Swing	Lab	4.58	4.19	46.5	NS	FORESIGHT	30	NS
	Con	1.82	1.75	50	NS	HINDSIGHT	28	NS
To Con 0.1-2%	Lab	5.70	5.31	43	NS	FORESIGHT	21	P<.025
	Con	3.06	4.42	28	NS	HINDSIGHT	40.5	NS
To Con 2-5%	Lab	4.14	4.63	42	NS	FORESIGHT	45.5	NS
	Con	5.06	4.44	46	NS	HINDSIGHT	41	NS
To Con 5-10%	Lab	1.68	2.94	20	P<.025	FORESIGHT	4	P<.01
	Con	4.98	5.13	46	NS	HINDSIGHT	17	P<.01
To Con > 10%	Lab	1.10	2.0	14.5	P<.01	FORESIGHT	15	P<.01
	Con	1.59	3.13	24	P<.05	HINDSIGHT	28.5	NS

Table 11.5 Summary of results for Q4 - Number of seats

OUTCOME	FORESIGHT-HINDSIGHT					LAB-CON		
	CONDITION	MEDIAN FORE	MEDIAN HIND	U	SIG	CONDITION	U	SIG
Lab Maj > 150	Lab	1.23	1.65	12.0	P<.01	FORESIGHT	42.5	NS
	Con	1.29	1.49	36.5	NS	HINDSIGHT	31.5	NS
Lab Maj 100-150	Lab	1.23	1.82	27	P<.05	FORESIGHT	42.5	NS
	Con	1.35	1.50	41	NS	HINDSIGHT	24	P<.05
Lab Maj 40-100	Lab	2.12	3.23	35.5	NS	FORESIGHT	38.5	NS
	Con	1.59	1.67	45	NS	HINDSIGHT	4.5	P<.01
Lab Maj 0-40	Lab	5.71	5.30	36	NS	FORESIGHT	12	P<.01
	Con	2.12	2.43	34.5	NS	HINDSIGHT	26	P<.05
Con Maj 0-40	Lab	5.31	4.53	42.5	NS	FORESIGHT	41	NS
	Con	4.62	4.82	47	NS	HINDSIGHT	49	NS
Con Maj 40-90	Lab	2.69	3.77	28	NS	FORESIGHT	19	P<.01
	Con	4.88	5.31	41	NS	HINDSIGHT	34.5	NS
Con Maj > 90	Lab	1.08	2.03	15	P<.01	FORESIGHT	17	P<.01
	Con	2.42	3.26	38	NS	HINDSIGHT	34.5	NS

Table 11.6 Summary of Results for Q5(a) - The Liberals % of the votes polled

OUTCOME	FORESIGHT-HINDSIGHT					LAB-CON		
	CONDITION	MEDIAN FORE	MEDIAN HIND	U	SIG	CONDITION	U	SIG
More 25%	Lab	1.16	1.73	19	P<.01	FORESIGHT	35	NS
	Con	1.71	1.50	41	NS	HINDSIGHT	22	P<.025
Between 19-25%	Lab	1.12	2.56	25	P<.05	FORESIGHT	26	P<.05
	Con	2.24	1.50	29.5	NS	HINDSIGHT	14	P<.01
Between 17-19%	Lab	1.59	4.18	25.5	P<.05	FORESIGHT	31	NS
	Con	3.24	2.02	42.5	NS	HINDSIGHT	35.3	NS
Between 12-17%	Lab	4.12	4.83	45.5	NS	FORESIGHT	25	P<.05
	Con	2.20	3.31	34.5	NS	HINDSIGHT	36	NS
Less 12%	Lab	4.36	3.36	49	NS	FORESIGHT	41.5	NS
	Con	3.0	4.07	43	NS	HINDSIGHT	50	NS

Table 11.7 Summary of Results for Q5(b) - The Liberals Swing to/from Liberal

OUTCOME	FORESIGHT-HINDSIGHT					LAB-CON		
	CONDITION	MEDIAN FORE	MEDIAN HIND	U	SIG	CONDITION	U	SIG
To Lib 5%	Lab	1.50	1.88	26	P<.05	FORESIGHT	38.5	NS
	Con	1.59	1.50	31	NS	HINDSIGHT	14	P<.01
To Lib 3-5%	Lab	1.70	2.50	21	P<.025	FORESIGHT	28	NS
	Con	2.22	1.73	27.5	NS	HINDSIGHT	22	P<.025
To Lib 0.1-3%	Lab	2.94	3.32	34	NS	FORESIGHT	45.5	NS
	Con	2.94	1.75	47.5	NS	HINDSIGHT	40.5	NS
No Swing	Lab	2.82	3.94	32	NS	FORESIGHT	40.5	NS
	Con	2.35	2.94	48	NS	HINDSIGHT	45	NS
Away Lib 0.1-3%	Lab	3.36	4.56	25	P<.05	FORESIGHT	46.5	NS
	Con	2.18	4.13	32	NS	HINDSIGHT	47	NS
Away Lib 3-5%	Lab	3.73	3.60	44	NS	FORESIGHT	34.5	NS
	Con	2.12	5.00	30	NS	HINDSIGHT	49.5	NS
Away Lib > 5%	Lab	3.59	2.56	40	NS	FORESIGHT	19	P<.01
	Con	1.59	3.09	20.5	P<.025	HINDSIGHT	47	NS

Table 11.8 Summary of Results for Q.5c - The Liberals Number of Seats

OUTCOME	FORESIGHT-HINDSIGHT					LAB-CON		
	CONDITION	MEDIAN FORE	MEDIAN HIND	U	SIG	CONDITION	U	SIG
More 30	Lab	1.19	1.69	20	P<.025	FORESIGHT	32.5	NS
	Con	1.59	1.50	45	NS	HINDSIGHT	33.5	NS
23-30	Lab	1.23	2.00	22	P<.025	FORESIGHT	26	P<.05
	Con	2.18	1.56	20	P<.025	HINDSIGHT	12	P<.01
14-23	Lab	1.53	3.02	25.5	P<.05	FORESIGHT	24.5	P<.05
	Con	3.71	1.63	21.0	P<.025	HINDSIGHT	20.5	P<.025
13	Lab	2.70	4.00	30.5	NS	FORESIGHT	35	NS
	Con	4.20	2.75	48	NS	HINDSIGHT	46	NS
6-12	Lab	4.29	4.31	48	NS	FORESIGHT	34.5	NS
	Con	3.35	3.50	39.5	NS	HINDSIGHT	36.5	NS
4-6	Lab	2.94	3.74	48	NS	FORESIGHT	35.5	NS
	Con	2.37	4.86	25	P<.05	HINDSIGHT	43	NS
Less 4	Lab	1.71	2.49	24	P<.05	FORESIGHT	43.5	NS
	Con	1.53	2.13	32	NS	HINDSIGHT	42	NS

Table 11.9 Summary of Results for Q6 - The National Front, % of poll

OUTCOME	FORESIGHT-HINDSIGHT					LAB-CON		
	CONDITION	MEDIAN FORE	MEDIAN HIND	U	SIG	CONDITION	U	SIG
Poll > 1%	Lab	1.35	1.88	28	NS	FORESIGHT	43.5	NS
	Con	1.78	1.88	37	NS	HINDSIGHT	42	NS
Poll 0.4-1%	Lab	4.13	3.81	47.5	NS	FORESIGHT	39	NS
	Con	4.83	5.25	46	NS	HINDSIGHT	41	NS
Poll 0.4%	Lab	4.47	4.56	40	NS	FORESIGHT	48	NS
	Con	5.12	5.19	48	NS	HINDSIGHT	39	NS
Poll 0.2-.4%	Lab	3.82	3.81	44	NS	FORESIGHT	30.5	NS
	Con	2.20	3.07	34.5	NS	HINDSIGHT	47	NS
Poll < 0.2%	Lab	2.47	2.13	40	NS	FORESIGHT	41	NS
	Con	1.76	2.50	25	P<.05	HINDSIGHT	44.5	NS

Table 11.10 Summary of Results for Q7 - The Scottish Nationalist Party

OUTCOME	FORESIGHT-HINDSIGHT					LAB-CON		
	CONDITION	MEDIAN FORE	MEDIAN HIND	U	SIG	CONDITION	U	SIG
More 20 seats	Lab	1.29	1.81	26	P<.05	FORESIGHT	49.5	NS
	Con	1.59	1.88	25	P<.05	HINDSIGHT	39	NS
13-20 seats	Lab	1.53	2.50	27	P<.05	FORESIGHT	33	NS
	Con	2.19	2.50	32.5	NS	HINDSIGHT	43	NS
10-12 seats	Lab	2.82	3.31	49	NS	FORESIGHT	39.5	NS
	Con	3.47	4.06	37	NS	HINDSIGHT	35.5	NS
6-10 seats	Lab	4.94	4.00	39	NS	FORESIGHT	48.5	NS
	Con	4.15	3.00	30	NS	HINDSIGHT	39.5	NS
Less 6 seats	Lab	4.71	4.63	42	NS	FORESIGHT	20	P<.025
	Con	2.47	1.88	29	NS	HINDSIGHT	23	P<.025

However, Conservative subjects saw this outcome as significantly more likely in hindsight. For outcomes which did not occur, the only significant difference (greater 78%: Conservative subjects) was in the opposite direction to that expected, i.e. foresight subjects saw this as less likely than hindsight subjects.

Q.3. The Swing

The outcome which occurred (swing to Conservative of between 5-10%) was seen as significantly more likely by Labour hindsight subjects than Conservative hindsight subjects, as Table 11.4 shows, but not by Conservative subjects in hindsight. For the outcomes which did not occur, the five significant differences shown in Table 11.4 are all in the opposite direction to that predicted by hindsight bias.

Q.4. Number of Seats

Table 11.5 shows that there were no significant differences between Labour and Conservative subjects in foresight and hindsight for the outcome which occurred (Conservative majority of 40-90 seats), although both hindsight medians are in the predicted direction. The three significant differences for outcomes which did not occur, shown in Table 11.5, are all in the opposite direction to that expected.

Q.5. The Liberals

(a) % of the votes polled. Table 11.6 shows foresight/hindsight comparisons to be in the direction expected for the outcome which occurred (between 12-17%), but not significantly so. Once again the significant differences for the outcomes which did not occur were all in the opposite direction.

(b)' Swing to/from Liberal. Table 11.7 shows for the outcome which occurred (swing away from Liberal of between 0.1-3%), the hypothesis to be supported for Labour subjects but not for Conservative subjects. Once again all significant differences between foresight/hindsight judgements for the outcomes which did not occur were in the opposite direction to that predicted by hindsight bias.

(c) Number of seats. Table 11.8 shows both hindsight medians to be in the expected direction for the outcome which occurred (between 6-12 seats), but not significantly so. The table also reveals seven hindsight assessments to be significantly differently for outcomes which did not occur. Two of these being in the expected direction (Conservative subjects: 23-30 seats; 14-23 seats), the other five in the opposite direction.

Q.6. The National Front

No significant differences were found for the outcome which occurred (poll between 0.4-1% votes), as Table 11.9 shows. The one significant difference for outcomes which did not occur was in the opposite direction to that expected.

Q.7. The Scottish Nationalist Party

For the outcome which occurred (less than 6 seats) there were no significant differences in foresight/hindsight, as Table 11.10 shows. The three significant differences for outcomes which did not occur were, once again, all in the opposite direction to that expected.

Summary '

There are ten outcomes, reflecting different aspects of the General Election, which occurred. These afforded twenty individual comparisons, half between Labour and half between Conservative subjects in foresight and hindsight. Fifteen of the hindsight medians supported the hypothesis, four significantly so. Five of the comparisons were non-hypothesis supporting, none significantly so.

From the seven questions, fifty two specified outcomes to the General Election did not occur, thus offering 104 individual comparisons in foresight and hindsight. 30 hindsight medians were in the predicted direction (i.e. judgements less in hindsight than foresight), but only 3 significantly so. 74 were in the opposite direction predicted by hindsight bias, 24 significantly so.

(b) Party Affiliations and judgements

Tables 11.1 to 11.10 summarise the medians given by Labour and Conservative party members. The tables also show the Mann Whitney U statistic for Labour/Conservative comparisons and whether the differences are reliable or not.

Q.1. The Result of the General Election

(a) Likelihood assessments. Table 11.1 shows that an overall majority to the Conservatives (the outcome which occurred) was perceived as significantly more likely by Labour subjects than Conservatives, in foresight, but not in hindsight. A labour majority was also perceived as more likely by Labour than Conservative, both in foresight and hindsight. No other significant differences for outcomes

which did not occur were found.

(b) Estimates of surprise. Table 11.2 shows that Labour subjects, in both foresight and hindsight, expressed significantly more surprise for the outcome which occurred than Conservative subjects. For outcomes which did not occur Conservative subjects expressed significantly more surprise than Labour to the outcomes: Labour majority, and overall Labour majority, both in foresight and hindsight.

Q.2. The turnout

In general high turnout is said to favour Labour. Labour subjects perceived a turnout of 73-78% (what happened) and greater than 78% as more likely in foresight, but not in hindsight, than Conservative subjects. There were no other significant differences.

Q.3. The Swing

The 5-10% swing to the Conservatives, which occurred, was seen as more likely by Conservative subjects in both foresight and hindsight. The various swings to Labour were, generally, seen as more likely by Labour subjects in hindsight but not foresight, as Table 11.4 shows. A swing to Conservative of between 0.1-2% was seen as more likely in foresight by Labour (such a swing would have kept Labour in office). A Conservative swing of greater than 10% was seen as more likely by Conservative subjects in foresight but not hindsight.

Q.4. The number of seats

Table 11.5 shows that a Conservative majority of between 40-90 seats (what happened) was seen as significantly more likely in foresight

by Conservative than Labour subjects. For other outcomes, those specifying a Labour majority of seats were seen as more likely by Labour. The outcome specifying a Conservative majority of greater than 90 seats was seen as more likely in foresight, but not in hindsight,by Conservative subjects.

Q.5. The Liberals

A Liberal gain in seats, per cent of votes polled or swing in their favour is generally regarded as favouring Labour. This is because additional votes for the Liberals tend to come from people who voted Conservative on previous occasions.

(a) % votes polled. As Table 11.6 shows, three of the four significant results - hindsight: more 25%, 19-25% and 12-17% - were seen as more likely by Labour subjects than Conservative. These are consistent with the above statement. One significant difference (foresight 19-25%, more likely by Conservative subjects), however, is not consistent with this.

(b) The Swing. Table 11.7 shows that the three significant results - swing to Liberal of more than 5% seen as more likely in hindsight by Labour; swing to Liberal of 3-5% more likely in hindsight by Labour; and swing away from Liberal of greater than 5% seen as more likely in foresight by Conservative - are also consistent with the above statement.

(c) The number of seats. Table 11.8 shows two significant results - 23 to 30 seats and 14 to 23 seats seen as more likely by

Labour - are consistent with the above claim. The other two - foresight: 23-30 seats and 14-23 seats, more likely by Conservative - are not.

Q.6. The National Front

The National Front poll such a small percentage of the vote that their fortunes are of little consequence to the two major parties. This is reflected in the results, shown in Table 11.9, where no significant differences were found between Labour and Conservative subjects, in either foresight or hindsight.

Q.7. The Scottish National Party

Table 11.10 shows that the only significant differences between Labour and Conservative subjects were for the outcome which occurred (less than 6 seats). In both foresight and hindsight this outcome was seen as more likely by Labour members. This is consistent with the view that the Scottish Nationalist Party has, in the past, taken votes from Labour, so a poor result for the S.N.P. would favour Labour. This is further substantiated by the fact that Scotland was the only part of the British Isles to show a slight swing to Labour.

11.4 DISCUSSION

The first issue this experiment was designed to explore was whether hindsight bias would be exhibited by subjects who had strong commitments, desires and attitudes for certain outcomes to come about.

Active Labour and Conservative party members given questionnaires about the General Election before and after the result was known

provided only partial and weak support for hindsight bias. This came exclusively from the outcomes which occurred to various aspects of the General Election. The twenty comparisons made for such outcomes showed 15 to support hindsight bias, with four of the differences being reliable ones. None of the five differences which were non-hypothesis supporting were reliable.

No support at all for hindsight bias was found for the outcomes which did not occur. In fact the large number of significant differences in the opposite direction to that predicted as compared with those in the predicted direction (27 and 3 respectively) might be seen as evidence for an opposite effect. That is, subjects perceived outcomes which did not happen as more likely in hindsight than foresight. This was found for only those questions where seven-point scales were used to measure uncertainty assessments.

These results, for the outcomes which did not occur, are difficult to explain: for one thing it is the first time such a pattern has been found in all the experiments reported so far in this thesis. Research by Fischhoff and his colleagues, as discussed in Chapter 3, and the experiments reported here appear consistently to show hindsight bias to be less pronounced for outcomes which did not happen or are not reported as occurring. However, none of Fischhoff's research or that reported here provides any indication of an effect in the opposite direction.

One possible explanation of these results is that subjects in hindsight are using the seven-point likelihood scales differently.

That is, those in hindsight are not using the "not very likely" pole of the scale in the same way as subjects in foresight. This can only remain a speculation as there is no other evidence for this and no reason, other than the pattern of results to regard this as the explanation.

The second issue this experiment was designed to explore concerned the effect of party affiliations and outcomes party members would have liked to have seen occur upon subjective likelihood assessments. It was predicted that outcomes favourable to a party would be seen as more likely, in both foresight and hindsight, by that party than the other one. The results offered strong support for this hypothesis. Generally, outcomes favourable to the Labour party were seen as more likely by Labour party members than Conservative party members. Conversely, outcomes favourable to Conservative were seen as more likely by Conservative than Labour party members.

The third issue was that hindsight bias would be in greater evidence for outcomes which happened in the party for which those outcomes were favourable. This received no support. The result of the General Election was highly favourable to Conservative: in foresight these outcomes were seen as highly likely by Conservative subjects. Hence, the extent of party affiliation influence upon subjective likelihood assessments was tending towards some ceiling before the outcome was known. For Labour party members a different pattern of results emerged for these outcomes in foresight and hindsight. In foresight, Labour subjects perceived these events as much less likely than Conservative subjects. Outcome knowledge may

have caused Labour subjects, but not Conservative, to realise that their assessments were unrealistically optimistic. In consequence, they adjusted them upwards to what they thought would have been more realistic at the time.

In general, these results show that subjective likelihood assessments are strongly influenced by the value of an outcome to subjects and their desire for it to occur. This positive relationship appears to be only marginally affected by outcome knowledge. On the basis of these findings it might be concluded that hindsight bias is much less likely to occur when concomitant attitudes and wishes to see desirable outcomes occur is high.

The nature of the task and the subjects used raises the question of what the subjective likelihood assessments were assessments of. Ideally, subjects are beings asked to give some formal expression to their true expectations, i.e. to put aside what they would like to see happen and objectively judge what they think will happen. However, it is apparent that the judgements reflect subjects' wishes. Hence in cases where a person has some kind of vested interest a bias as serious as that due to hindsight may occur. People who see undesirable outcomes as less likely, and desirable ones as more likely, than they are and fail to modify what they thought once they know how events turned out are going to learn even less from the past than if they were subject to hindsight bias.

In conclusion, this experiment has demonstrated that material relevant to subjects is necessary in order to discover how widespread

a phenomena hindsight bias is in "everyday life". The experiment has also extended the findings of the previous experiment by showing how subjects' desire and wish to see certain outcomes obtain and othersnot influence subjective likelihoods and interact with outcome knowledge.

CHAPTER 12

EXPERIMENT 8[1]: ROLES OF WOMEN IN SOCIETY: ESTIMATES IN FORESIGHT AND HINDSIGHT

12.1. INTRODUCTION

The final experiment to be reported in this thesis investigates foresight/hindsight judgements using factual material. In Chapter 3, it will be recalled, two studies were reviewed, Fischhoff (1977) and Wood (1978), which also used such stimulus material. In that chapter it was shown that those experiments provided the strongest evidence, relative to those which used "laboratory material" or contemporary social events, for hindsight bias. The experiment to be reported here seeks to replicate and extend these findings.

Previous research using factual stimulus material to investigate the effect of outcome knowledge upon judgements has asked people to assess the likelihood that a particular definition of a word is correct (or incorrect). In contrast to this technique the experiment to be reported here used statistical information concerning the number or percentage of women in certain careers or roles in society. In consequence, the use of such material allows us to ask subjects to estimate what they think or thought such numbers or percentages are or were.

---

1. Lorraine Sherr assisted in the collection of datal for this experiment, she also provided the author with the experimental material.

With an objective criteria by which to compare estimates made in foresight and hindsight it was predicted that those who were told what the actual figures were (hindsight) would produce estimates closer to those actual figures than those who were not told the figures (foresight).

12.2. METHOD

SUBJECTS

87 subjects, 25 male and 52 female, attending an Open Day at the Department of Psychology, University of Warwick, took part in the experiment.

PROCEDURE

Subjects were run in small groups, and randomly assigned to either a foresight or hindsight condition. In both conditions subjects were required to estimate the number or percentage, as appropriate, of women in certain professions or roles in society. There were sixteen questions in all, ten relating to women in employment at the University of Warwick, and six concerned with the percentage of women occupying certain roles in England.

Subjects in the foresight condition were merely asked to estimate the number or percentage of women in those jobs or roles. Subjects in the hindsight condition were provided with the actual statistics in one column on the questionnaire and asked to estimate what they thought the figures might have been.

The entirely random procedure of assigning subjects to one of the two conditions led to 13 and 12 males in foresight and hindsight

respectively, and 26 and 36 females in foresight and hindsight respectively. Subjects spent about 5-10 minutes answering the questionnaire.

INSTRUCTIONS

Subjects in the foresight condition received the following instructions to the first 10 questions:

> At the University of Warwick there are the following numbers of people. How many do you think are women?

For the final six questions concerned with the percentage of women in certain roles in England no further instructions were given as it was thought that the questions themselves were self-explanatory.

Subjects in hindsight received the following instructions for all sixteen questions:

> For the questionnaire below there are a series of questions followed by two columns (labelled 1 and 2). Column 2 refers to the actual figures. What you are asked to do is estimate what you thought the figures might have been. Do this in Column 1.

Subjects then received similar instructions as those in foresight to the first 10 questions except that the second sentence was modified to read as follows:

> Your response should estimate the number of women.

QUESTIONNAIRE

The sixteen questions used in this experiment were taken from

"The Report of the Sex Equality Committee of Senate", University of Warwick, 1979. Subjects in both conditions received the same 16 questions. Reproduced below is the questionnaire given to hindsight subjects, foresight subjects received the same questionnaire but without column 2 (outcome knowledge).

	QUESTION	COLUMN 1	COLUMN 2
1.	62 Professors	---	1 is a woman
2.	65 Readers and Senior Lecturers	---	4 are women
3.	261 Lecturers	---	29 are women
4.	16 Registry Senior Staff	---	3 are women
5.	7 Senior Library Staff	---	0 are women
6.	18½ Library Assistants (½ = part-time)	---	17½ are women
7.	Secretaries	---	100% are women
8.	Domestic Cleaners, Cooking Assistants	---	100% are women
9.	In the English Department last year 71% of the undergraduate degrees were obtained by women. 46 postgraduates were admitted, what percentage were women?	---	50% were women
10.	Warwick has (last year) 355 full-time post-graduates, how many women?	---	110 were women
	* * * * * *		
11.	What percentage of the labour force in England are women?	---	38%
12.	How many of these are also Housewives?	---	75%
13.	In the nursing profession, what percentage are women	---	90%
14.	Mothers of 0-2 year olds, how many are at home?	---	85%
15.	Mothers of 5-15 year olds, how many are at home?	---	56%
16.	What percentage of part-time workers are women?	---	88%

12.3 RESULTS

To discover how outcome knowledge affected estimates two sets of analyses between conditions were made. First, in order to get a general picture of the results the means between foresight and hindsight, summarised in Table 12.1, were compared. If the mean in the hindsight condition for a particular question was closer to the actual figure (shown in column 2 of Table 12.1) it was regarded as hypothesis supporting. If, on the other hand, the mean in hindsight was less accurate than the one in foresight it was regarded as non-hypothesis supporting. This procedure was followed for both males and females.

For females the number of hypothesis - supporting hindsight means was 15, with one being non-hypothesis supporting. Similarly, for males the figures were 15 hypothesis-supporting and one nonhypothesis-supporting.

This general description fails to take account of the magnitude of the difference between the two conditions for each question. In order to get at this the second set of analyses compared estimates given in foresight and hindsight to each question by analysis of variance. The factors foresight/hindsight and male/female were extracted, giving a 2 x 2 independent analysis with unequal cells (Winer, 1971, p. 445). Table 12.2 summarises the results of this analysis for each of the 16 questions.

The ten questions concerned with the number or percentage of women in certain types of employment at the University of Warwick yielded seven significant differences between foresight and hindsight

12.3 RESULTS

To discover how outcome knowledge affected estimates two sets of analyses between conditions were made. First, in order to get a general picture of the results the means between foresight and hindsight, summarised in Table 12.1, were compared. If the mean in the hindsight condition for a particular question was closer to the actual figure (shown in column 2 of Table 12.1) it was regarded as hypothesis supporting. If, on the other hand, the mean in hindsight was less accurate than the one in foresight it was regarded as non-hypothesis supporting. This procedure was followed for both males and females.

For females the number of hypothesis - supporting hindsight means was 15, with one being non-hypothesis supporting. Similarly, for males the figures were 15 hypothesis-supporting and one nonhypothesis-supporting.

This general description fails to take account of the magnitude of the difference between the two conditions for each question. In order to get at this the second set of analyses compared estimates given in foresight and hindsight to each question by analysis of variance. The factors foresight/hindsight and male/female were extracted, giving a 2 x 2 independent analysis with unequal cells (Winer, 1971, p. 445). Table 12.2 summarises the results of this analysis for each of the 16 questions.

The ten questions concerned with the number or percentage of women in certain types of employment at the University of Warwick yielded seven significant differences between foresight and hindsight

Table 12.1 Summary of Mean Estimates and Standard Deviations given in Foresight and Hindsight by Males and Females

QUESTION	ACTUAL FIGURE	ESTIMATED FIGURES							
		FORESIGHT MALES		FORESIGHT FEMALES		HINDSIGHT MALES		HINDSIGHT FEMALES	
		$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
1. 62 Professors	1	17.31	9.10	19.56	6.28	10.66	5.35	15.67	9.45
2. 65 Readers and Senior Lecturers	4	23.0	8.20	22.96	8.15	15.17	7.28	17.41	9.85
3. 261 Lecturers	29	109.9	49.1	45.9	19.2	75.4	41.7	69.0	31.3
4. 16 Registry Staff	3	9.23	2.62	7.81	2.73	4.42	2.71	6.11	2.74
5. 7 Senior Library Staff	0	4.54	0.88	4.08	1.41	3.75	1.54	3.41	1.54
6. 18.5 Library Assistants	17.5	13.9	2.56	12.75	3.15	13.6	5.34	13.1	4.15
7. Secretaries (%)	100	87.0	8.56	89.8	15.11	96.3	8.29	95.3	11.93
8. Domestic Cleaner, Cooking Assistants	100	89.6	10.15	86.8	12.05	92.8	10.27	96.1	9.47
9. % Women Postgraduates in the English Department	50	60.0	16.2	49.1	17.23	52.6	13.85	51.3	15.83
10 Total (355) postgraduate women in whole University	110	134.0	23.68	152.6	41.3	133.7	40.8	132.0	32.2
11. % Labour for women in England	38	33.4	7.0	39.0	11.0	36.1	6.36	36.0	7.64
12 % Also housewives	75	64.3	28.4	55.1	25.29	71.4	12.06	70.6	20.1
13 % Nurses in Nursing Profession	90	83.1	11.46	79.9	10.76	83.2	8.81	85.3	9.81
14 % mothers of 0-2 yr old at home	85	85.5	12.40	81.1	13.03	84.6	9.41	83.3	10.21
15 % mothers of 5-15 yr old at home	56	61.1	16.10	47.9	15.9	52.9	16.44	48.8	13.08
16 % part-time workers female	88	68.5	22.3	74.5	15.33	78.3	16.0	77.9	10.84

Table 12.2 Summary of 2 x 2 Analyses of Variance performed on each of the 16 Questions

SOURCE	SS	DF	MS	F	P
Q.1. THE PROFESSORS					
A (Foresight/Hindsight)	486.7	1	486.7	7.29	P < 0.01
B (Male/Female)	230.7	1	230.7	3.47	NS
A x B	33.1	1	33.1	0.49	NS
ERROR	5471.6	82	66.7		
Q.2. READERS AND SENIOR LECTURERS					
A	785.1	1	785.1	9.95	P < 0.005
B	21.3	1	21.3	0.27	NS
A x B	22.9	1	22.9	0.29	NS
ERROR	6394.9	82	78.95		
Q.3. LECTURERS					
A	78648.4	1	78648.4	9.83	P < 0.005
B	23468.8	1	23468.8	2.93	NS
A x B	27996.1	1	27996.1	3.49	NS
ERROR	664656.9	83	8007.9		
Q.4. REGISTRY SENIOR STAFF					
A	187.15	1	187.15	25.35	P < 0.001
B	0.32	1	0.32	0.04	NS
A x B	42.91	1	42.91	5.81	P < 0.025
ERROR	612.82	83	7.38		
Q.5. SENIOR LIBRARY STAFF					
A	9.06	1	9.06	4.53	P < 0.05
B	2.66	1	2.66	1.33	NS
A x B	0.08	1	0.08	0.41	NS
ERROR	161.89	81	2.00		
Q.6. LIBRARY ASSISTANTS					
A	928.7	1	928.7	0.87	NS
B	2670.8	1	2670.8	2.51	NS
A x B	15.89	1	15.89	0.02	NS
ERROR	87184.9	82	1063.2		
Q.7. SECRETARIES					
A	858.6	1	858.6	5.79	P < 0.025
B	12.31	1	12.31	0.08	NS
A x B	59.18	1	59.18	0.40	NS
ERROR	11427.2	77	148.51		
Q.8. DOMESTIC CLEANERS, COOKING ASSISTANTS					
A	617.9	1	617.9	5.63	P < 0.025
B	0.68	1	0.68	0.01	NS
A x B	143.9	1	143.9	1.31	NS
ERROR	8565.9	78	109.8		

Table 12.2 continued

SOURCE	SS	DF	MS	F	P
Q.9. WOMEN POSTGRADUATES IN ENGLISH DEPARTMENT					
A	119.9	1	119.9	0.46	NS
B	640.3	1	640.3	2.48	NS
A x B	395.1	1	395.1	1.53	NS
ERROR	20676.7	80	258.5		
Q.10. POSTGRADUATE WOMEN IN THE UNIVERSITY					
A	1838.6	1	1838.6	1.50	NS
B	1196.7	1	1196.7	0.97	NS
A x B	1725.3	1	1725.3	1.40	NS
ERROR	93376.0	76	1228.6		
Q.11. LABOUR FORCE IN ENGLAND WOMEN					
A	0.51	1	0.51	0.01	NS
B	137.0	1	137.0	1.86	NS
A x B	145.3	1	145.3	2.00	NS
ERROR	6048.9	83	73.48		
Q.12. ALSO HOUSEWIVES					
A	2147.1	1	2147.1	4.28	P < 0.05
B	424.1	1	424.1	0.85	NS
A x B	298.0	1	298.0	0.59	NS
ERROR	40608.8	81	501.3		
Q.13. NURSES IN NURSING PROFESSION					
A	108.9	1	108.9	1.21	NS
B	12.39	1	12.39	0.14	NS
A x B	101.2	1	101.2	1.12	NS
ERROR	7475.2	83	90.1		
Q.14. % OF MOTHERS WITH 0-2 YEAR OLD CHILDREN AT HOME					
A	7.48	1	7.48	0.58	NS
B	144.03	1	144.03	1.12	NS
A x B	45.54	1	45.54	0.35	NS
ERROR	10711.9	83	129.06		
Q.15. % OF MOTHERS WITH 5-15 YEAR OLD CHILDREN AT HOME					
A	233.2	1	233.2	1.05	NS
B	1326.2	1	1326.2	5.98	P < 0.025
A x B	361.9	1	361.9	1.63	NS
ERROR	18396.7	83	221.7		
Q.16. % PART-TIME WORKERS FEMALE					
A	775.1	1	775.1	3.43	NS
B	141.5	1	141.5	0.63	NS
A x B	186.2	1	186.2	0.82	NS
ERROR	18783.1	83	226.3		

estimates. These were all in the predicted direction. The questions which yielded these differences were as follows: Q.1. Professors ($F = 7.29$; $df = 1,82$; $P < 0.01$); Q.2. Readers and senior lecturers ($F = 9.95$; $df = 1,82$; $P < 0.005$); Q.3. Lecturers ($F = 9.83$; $df = 1,83$; $P < 0.005$); Q.4. Registry senior staff ($F = 25.35$; $df = 1,83$; $P < 0.001$); Q.7. Secretaries ($F = 5.79$; $df = 1,77$; $P < 0.025$); Q.8. Domestic cleaners and cooking assistants ($F = 5.63$; $df = 1,78$; $P < 0.025$).

The six questions concerned with the percentage of women occupying certain roles in society in England yielded only one significant difference (Q.15 Percentage of mothers of 5-15 year olds at home: $F = 4.28$; $df = 1,81$; $P < 0.05$). This was also in the predicted direction.

The analysis also allowed us to look for sex differences. Only one question, Q.15. percent of mothers of 5-15 year olds at home, yielded a significant difference ($F = 5.98$; $df = 1,83$; $P < 0.025$). Here females underestimated the percentage of women at home to a greater extent than males. There was one significant interaction (Q.4. Senior Registry Staff: $F = 5.81$; $df = 1,83$; $P < 0.025$). Here males showed little difference between foresight and hindsight estimates, whereas females gave more accurate estimates in hindsight than foresight.

12.4 DISCUSSION

This experiment was designed to test the hypothesis that those who were told the actual number or percentage of women in certain types of employment or roles in society would give estimates closer

to the true figure than those who were not so informed. The experiment was an investigation of hindsight bias because subjects were asked, in hindsight, to give the estimates they thought they would have had they not been told.

In general, the results provide strong support for hindsight bias. Of the sixteen hindsight estimates, fifteen were in the predicted direction, eight of these being significantly different to foresight estimates.

The first ten questions subjects gave estimates for, concerned with women in employment at the University of Warwick, produced seven significant differences, all in the predicted direction. The six questions concerned with women's roles in England produced only one significant difference, this in the predicted direction.

The lack of strong support for hindsight bias for these six questions can be attributed to the accuracy with which subjects in the foresight condition made estimates. This experiment, in contrast to others using factual material (i.e. Fischhoff, 1977; Wood, 1978) had objective criteria with which to compare subjects' estimates. One consequence of this is that if subjects in foresight make reasonably accurate estimates then there is hardly likely to be any effect due to giving outcome knowledge.

This is what has happened in Questions 11-16. Inspection of Table 12.1 reveals that, for these questions, the difference between the mean estimates in foresight and the actual figures is, generally

speaking, quite small. For example, with Question 14 it is 0.5 and 3.9 for males and females respectively. If we now turn to look at differences between hindsight estimates and the actual figures we again find them to be small. For Question 14, again, the differences are 0.4 and 1.7 for males and females respectively.

By way of contrast, the differences between the actual figures and mean estimates in foresight for the first 10 questions is quite large. The difference between the actual figures and hindsight mean estimates is substantially smaller.

Hence, we can only expect hindsight bias to occur when there are objective criteria with which to compare estimates if subjects in foresight make relatively inaccurate estimates. That there was such a "ceiling effect" operating here is lent further support by subjects' comments. Those in foresight, on the whole, expressed considerable surprise at the actual figures for the first 10 questions when they were told them after completing the experiment. However, they did not express such surprise when they were informed of the actual percentages for the final 6 questions.

One advantage of using material with an objective criteria, then, is that we can determine how accurately people make estimates in foresight and hindsight. Asking people to make subjective likelihood assessments, on the other hand, does not allow this kind of analysis. However, if we are more interested in subjects' judgements and perceptions of social events it is not easy, if at all possible, to obtain any indication of accuracy. For example, somebody might say that there

is a 70% chance of an all-out nuclear war between American and the U.S.S.R. in the next five years. It is not clear how one would determine the accuracy of this judgement. At most, we can only say that the person was wrong if they had said it was 100% likely and the halocaust did not occur, or if the person said it was 0% likely (i.e. no chance at all) and there was a nuclear war.

In conclusion, this experiment has replicated and extended the findings of Fischhoff (1977) and Wood (1978). The experiment has provided strong evidence for hindsight bias where factual stimulus material is used. The experiment has extended the findings as it has found the bias to occur when people make estimates of actual frequencies rather than subjective likelihood assessments of the chances of being right or wrong about some question of fact.

PART III

GENERAL DISCUSSION AND THEORETICAL CONSIDERATIONS

CHAPTER 13

HINDSIGHT BIAS AND ORDER EFFECTS

13.1 INTRODUCTION

This chapter considers two main issues to arise out of the experiments reported in Part II of this thesis. Firstly, a discussion of hindsight bias. This is considered specifically from the point of view of the three types of experimental material - laboratory, contemporary social/political events, and factual - used to investigate the effect of outcome knowledge upon perceptions and subjective likelihoods. It is followed by a general discussion of the results. Secondly, the effect of order of information upon judgements is discussed.

13.2 HINDSIGHT BIAS

The experiments reported in PART II, and the previous research reviewed in Chapter 3, of this thesis investigated hindsight judgements using three types of experimental material. This section discusses these by following the same format and provides some indications of the conditions under which hindsight bias may be expected to occur and the conditions under which it may not occur.

13.2.1 Experiments using laboratory material

Experiments 1 to 3 (reported in Chapters 5 to 7 respectively) investigated hindsight/foresight judgements using "laboratory material", as characterised in Chapter 3. Experiments 1 and 3, it will be recalled, investigated hindsight bias, Experiment 2 creeping determinism.

Experiments 1 and 3 did not, in general, provide strong support for hindsight bias. Both failed to produce a systematic pattern of findings in this respect. What support there was for the claim that outcome knowledge leads to bias indicated that if it is going to occur at all it does so as an event description unfolds and is not present immediately people are in possession of such knowledge.

Furthermore, the bias may be influenced by order of information. A consistent, but weak pattern of findings, in both experiments indicated that information bearing on the outcome reported to have occurred may encourage the bias when it appears at the end of the scenario, i.e. a recency effect. Conversely, information indicating outcomes other than the one reported may, when it appears at the end of the scenario, reduce the bias, again a recency effect. It must be stressed that the above findings were not particularly convincing, being

less so in Experiment 3 than Experiment 1. Some reasons for this were discussed in Chapter 7 and will not be repeated here.

Experiments 1 and 2, investigating hindsight bias and creeping determinism respectively, both indicated, although again not particularly convincingly, that hindsight effects may not be so marked when outcome knowledge is given at the beginning, rather than the end, of the event description. Experiment 1 provided stronger support for this than Experiment 2.

All three experiments demonstrated, in keeping with previous work by Fischhoff and his colleagues, that hindsight effects are less frequent for outcomes believed not to have happened than ones reported as occurring. It is of interest to speculate why this is the case, but as it is a finding which occurs with experiments using different types of material it will be considered in the final part of this section.

Overall, the experiments using laboratory material do not provide strong support for Fischhoff's claim that people in possession of outcome knowledge exhibit hindsight bias or creeping determinism in judgements under uncertainty. The experiments indicate that when bias occurs it only does so as the event description unfolds; may be removed or reduced by varying the order of information and by reporting outcomes before rather than after (as Fischhoff, 1975a did) the scenario.

13.2.2 Experiments using contemporary Socio-Political Events

Experiments 5 to 7 (reported in Chapters 9 to 11 respectively) all investigated hindsight bias using contemporary social or socio-political events. All experiments offered some support for the bias and also indicated conditions which encourage or reduce it.

The two experiments (5 and 7) which investigated hindsight judgements with respect to socio-political events (the British Firemen's Strike of 1977/78 and the General Election of May, 1979), taken together, failed to provide consistent support for hindsight bias. Whilst Experiment 5 did provide some strong support, Experiment 7 showed that when the stimulus material is highly relevant to subjects and subjects have strong desires, commitments, etc. to see certain outcomes about hindsight bias is unlikely to occur.

Experiment 6, in contrast, showed that designing an experiment on the grounds of relevance alone is insufficient to reduce or eliminate hindsight bias. In this experiment, it will be recalled, female subjects made judgements about what they thought their chances of being pregnant were. Here hindsight bias was found when women had a positive pregnancy test result but not when they had a negative one.

Both Experiments 5 and 6 demonstrated that subjects who made both foresight and hindsight judgements showed less bias than subjects who made only hindsight assessments. This, as mentioned in Chapter 9, confirmed previous findings by Fischhoff and his colleagues.

Experiment 5 also showed two other things to affect the degree of influence of outcome knowledge upon judgements. First, bias is more likely where people are asked only to assess the likelihood of outcomes presented to them rather than if they are asked both to produce their own outcomes and assess the likelihood of each. Second, certain types of information were found to affect the bias. Antecedent events were more likely to result in bias than events surrounding the outcome which happened. This was an unexpected finding and some theoretical discussion of it was given in Chapter 9.

Using a general conception of attitude, Experiment 7, as indicated above, clearly showed how this affected judgements. Experiment 5 also attempted to investigate this. However, as discussed in Chapter 9, this experiment failed to show any effect of attitudes upon subjective likelihood assessments. This being mainly due, as argued there, to the lack of any explicit experimental manipulation of attitudes. However, both experiments demonstrated a lack of interaction between attitudes, desires, etc. and outcome knowledge to make hindsight bias stronger under certain circumstances.

In general, the experiments using contemporary social or socio-political material provided important insights into the extent of hindsight bias we might expect in the "real world" or everyday life. They also provided indications of some of the conditions under which bias is likely and not likely to occur.

13.2.3 Experiments using factual material

The only experiment reported in this thesis to investigate hindsight bias using factual material was Experiment 8. This differed from Fischhoff (1977) and Wood (1978) because it asked people to estimate actual numbers or percentages rather than make subjective likelihood assessments.

The experiment provided strong support for hindsight bias. The majority of estimates made to the questions demonstrated that subjects who knew what the actual figures were gave more accurate estimates, than those who were not in possession of such knowledge.

Whilst the experiment does give strong support for the bias a problem arises over whether the results would generalise to subjective

likelihood assessments made with the same material. In view of the fact that the other experiments reported in this thesis did not produce such a systematic or convincing set of results as those claimed by Fischhoff such a generalisation might be problematic. Obviously, this is a line of enquiry which demands further empirical investigation.

13.2.4 Summary

The seven experiments investigating hindsight judgements, six looking at hindsight bias and one at creeping determinism, provide varying degrees, ranging from weak to strong, of support for previous research by Fischhoff and his colleagues.

The weakest support comes from that using laboratory material and the strongest from that using factual material. Those experiments investigating such judgements in an "applied" setting provided intermediate support relative to experiments using other types of stimulus material. These "applied" experiments also provided the best indications of conditions likely to encourage, reduce or eliminate hindsight bias. Most notable of these were the findings that asking people to produce their own outcomes reduced bias and providing antecedent details increased it. Also, people who wanted to see certain outcomes happen showed little bias. Here, though, attitudes, desires, etc., had such a strong influence on subjective likelihood assessments that a different bias to that of hindsight was shown to be present.

Generally, throughout the experiments it was found that hindsight bias/creeping determinism is more likely to be present, if at all, for the outcome which occurred rather than for the outcomes which did not occur. The question arises as to why this might be the case (an explanation

of hindsight bias itself is treated in a later section of this chapter).

Fischhoff and Beyth (1975a), Fischhoff (1977) offer an explanation of this differential hindsight effect by considering the different information processing demands involved in each circumstance. Their arguments will not be repeated here, the reader is referred back to Chapter 3, except to add that they are lent further credence in light of the "confirmatory hypothesis - testing strategy" people use in predictive tasks discussed in Chapter 2. Besides such a plausible theoretical explanation an obvious methodological one needs to be considered.

In all the experiments which have used social scenarios or socio-political events, subjects have been provided, typically, with four or more outcomes to some event. One of these is reported to subjects as having occurred or did in fact, occur. Hindsight bias for outcomes not reported to have occurred or which did not occur may not be detected simply because the reduced likelihood is spread over three or more such outcomes. This means that whilst the increase in likelihood given to the outcome which occurred often results in a significant difference the corresponding decrease is spread over the three or more outcomes which did not occur. This results in individual differences unlikely to be detected. One way to test this would be to provide a smaller number of outcomes to a scenario, for example three, and see if hindsight bias is found for the two outcomes not reported to have occurred.

On the face of it Experiment 6 might be thought to throw light on this issue also because only one of two outcomes were possible. However, because subjects only gave one assessment on a scale which covered both eventualities (positive or negative result to the pregnancy test) it cannot help here.

Further research, along the lines indicated above, is required to determine whether a theoretical explanation is needed or whether the result is an artifact of the design of the experiments.

13.3 ORDER EFFECTS

Five of the eight experiments (Experiments 1, 2, 3, 4 and 5) reported in this thesis investigated the effects of different information orders upon judgement under uncertainty. Four of these experiments (Experiments 1, 2, 3 and 5) looked at this with respect to foresight/hindsight judges and one (Experiment 4) investigated just order effects. This section will briefly summarise the findings and draw some conclusions about how order of information may be expected to influence judgements generally.

The three experiments using laboratory material to investigate the effect of different orders of information upon hindsight judgements provided only partial evidence for recency effects. The strongest evidence came from Experiment 1 where greater hindsight bias was found when statements indicating the reported outcome appeared at the end of the event description rather than at the beginning. Some evidence for recency effects was found for outcomes reported not to have happened as well.

The next two experiments, however, produced little evidence for order effects in this context. Experiment 3, which was designed specifically with information order in mind, failed to produce any reliable results to indicate that hindsight bias might be affected by order of information. In this experiment, the medians were generally, in the predicted direction, but not significantly so. Some methodological/design reasons explaining why this might have occurred were given in

Chapter 7 and will not be repeated here.

One tentative conclusion that we might draw from these experiments is that hindsight bias may be influenced by information order when it is presented sequentially (as in Experiment 1) but not when it is presented all at once, as in the continuous prose passages in Experiments 2 and 3. This might be taken as an indication that historians, policy-makers, etc. who receive or search for information one item at a time may be more susceptible to hindsight bias if they come across that information in a certain order than if that information was received or searched for in a piecemeal fashion. However, further research is required before we could feel confident in accepting such a conclusion.

In a different vein, Experiment 5 demonstrated that ordering of different types of information produced a consistent and reliable influence on hindsight judgements. The experiment showed that antecedent details of an event, when presented after a summary of how the outcome to the event came about, encouraged hindsight bias.

It might be inferred, therefore, that possession of outcome knowledge leads people to perceive certain antecedent details (specifically, those indicating the outcome) as more salient than others. The results suggest that assimilation of this material is guided by possession of outcome knowledge also. This together with the research, discussed in Chapter 2, that people operate using a confirmatory hypothesis-testing strategy would suggest that information indicating the outcome which occurred is paid the most attention and hence more readily assimilated into memory.

Subsequent retrieval of information for the purposes of making judgements consists almost exclusively of information indicating the outcome, with the consequence that that outcome is seen as more likley in hindsight than foresight.

By contrast, people in foresight will not be subject to biased information assimilation and recall of that information simply because they do not know what has happened. Subsequent recall of information should be more representative of the degree of uncertainty at the time of the original discussion.

Results from Experiment 5 concerning the effect of different types of information have provided some important and interesting findings concerning the effect of order of information. The theoretical discussion offered above suggests a number of important lines of enquiry for future research. More general discussion will be given to these theoretical aspects when we come to consider all the experiments reported in this thesis in the next section.

Experiment 4 investigated witness and testimony order with respect to jurors' judgements of guilt or innocence in a rape trial case. As will be recalled, no order effects were found. An extended discussion attempting to account for these lack of findings was given in Chapter 8 and will not be repeated here. On the basis of that discussion we might tentatively conclude that order of information is not a major variable in psycho-legal contexts where trials last, typically, hours or days and order is manipulated within the restraints of the traditional structure of a criminal trial.

To summarise, the various manipulations of information order in the experiments reported in PART II of this thesis have provided mixed evidence as to the importance of such a variable. Overall, only two of

the five experiments demonstrate consistent order effects. One indicating recency and the other showing how different types of information affect judgements. With respect to hindsight bias order of information would seem to be an important variable when information is presented sequentially and when different types of information are under consideration.

CHAPTER 14

THEORETICAL CONSIDERATIONS

14.1 INTRODUCTION

The final chapter of this thesis provides a theoretical discussion of the cognitive processes, centred around Tversky and Kahneman's (1974) heuristics of thinking, that may explain hindsight bias. The chapter and thesis concludes by making suggestions for future research on hindsight judgements specifically, and that premised on a model of man as an intuitive psychologist generally.

14.2 Hindsight Bias and Cognitive Processes

This section discusses some theoretical implications of the findings from the experiments concerned with foresight/hindsight judgements. Some explanation is offered as to how hindsight bias arises by recourse to the heuristics of thinking and cognitive strategies people use which were discussed in Chapter 2. At the same time some aspects of memory, such as assimilation and accessibility of information are also discussed.

In Chapter 2 it was indicated that hindsight bias may come about as a result of the inappropriate use of certain cognitive processes or

heuristics of thinking. The experiments reported in Part II of this thesis provide some answer to the question of under what experimental condition heuristics produce biased judgements.

Most previous research by Tversky and Kahneman, Fischhoff and others has been concerned with establishing the existence of heuristics and biases. The research reported in this thesis has gone beyond this and allows us to give more precise answers about when use of certain heuristics will lead to bias.

The three heuristics of thinking (availability, representativeness and anchoring and adjustment) proposed by Tversky and Kahneman (1974) can all be seen to play a role in producing hindsight bias. Availability and representativeness, described in Chapter 2, are most appropriate to explain hindsight bias where it is demonstrated by subjects giving likelihood assessments. Anchoring and adjustment provides an explanation of hindsight bias when subjects make estimates, as in Experiment 8. This will be discussed later.

The availability heuristic (the ease with which relevant instances or occurrences can be brought to mind when assessing the likelihood of an event) is, on the surface, the most likely candidate for explaining hindsight bias. However, some researchers (e.g. Nisbett and Ross, 1980; Schneider, Hastorf and Ellsworth, 1979) appear to ignore the interdependence of availability and representativeness.

In what follows we shall (1) indicate how availability explains hindsight bias with reference to certain experimental findings reported in this thesis; (2) show how anchoring and adjustment explains hindsight bias in Experiment 8; (3) demonstrate how availability can also be used

to explain why the bias might not appear in the first place under certain conditions; and (4) discuss the interdependence of availability and representativeness.

14.2.1 Availability and Hindsight Bias

The general reason for regarding the availability heuristic responsible for hindsight bias is that once an outcome to some social event is known it becomes more available in memory than other information surrounding the event. In effect it is a highly salient piece of information, which as will be shown later, highlights the role of representativeness. Outcomes which did not occur are not so available, at a later time. The net effect, then, is that the perceived likelihood of the outcome which occurred increases in hindsight.

Specific reasons for using availability to explain hindsight bias derive from consideration of some of the experimental conditions under which the bias was found to occur.

Consider, for example, the finding in Experiment 6 that women who received a positive result to a pregnancy test exhibited bias, whilst those who received a negative result did not.[1] The reason for considering availability important here stems from the further finding that there was a trend showing positive/negative outcome to be related to whether contraception or no contraception (plus contraceptive failure) was used. One strategy women might have used in assessing how pregnancy could occur may depend on the construction of relevant scenarios or "scripts" (Abelson, 1976). The availability of such

---

[1] Incidentally, this pattern of results provides some evidence against the anchoring and adjustment heuristic being responsible for hindsight bias here. Briefly, for such an heuristic to be operating we would expect to find a consistent effect of outcome knowledge on both positive and negative results. This heuristic can offer us no plausible explanation of why one group should show the bias and the other not, hence it may be regarded as not operating here.

scenarios, the ease with which they can be brought to mind, will influence assessments. Hence women who did not use contraceptives or had contraceptive failure may realise that there is a higher chance of pregnancy than those who did use contraception simply because relevant scenarios indicating the outcome are relatively easy to imagine for this group.[2]

The negative group showed no bias because attendance at the clinic served to counteract the effect of outcome knowledge. Availability also plays a role here, this is because it is reasonable to assume that this group constructed other scenarios which provided them with the suspicion of being pregnant.

Another example, this time from Experiment 5, further demonstrates the role of availability. In this experiment one of the findings was that antecedent details concerning the first four weeks of the Firemen's Strike were found to encourage hindsight bias. Outcome knowledge, as discussed earlier in this chapter, increases the salience of those antecedent events which indicate the outcome which happened would happen. Furthermore, as Fischhoff (1975a) has shown, such knowledge also determines which items of information in an event description are perceived as relevant to understanding why the outcome happened.

This, together with a "confirmatory bias" in information search and recall (Snyder, in Press), will have the net effect of making such information the most "available". This will result in a relative ease in constructing scenarios which indicate the outcome which happened, and a relative difficulty in imagining why and how other outcomes which did not happen could have happened.

---

[2] In giving this explanation one would not want to ignore, of course, the plausible possibility that women whose test proved positive were aware of various psysiological cues that women whose test proved negative would have been unlikely to experience.

Whilst this discussion rests upon differential "accessability" of information, it is reasonable to argue that outcome knowledge endows certain antecedent information with greater subjective accessibility because it is made more salient. The consequence being to make this information more "available" from memory.

Experiment 1 provides a final example of how availability accounts for hindsight bias. In this experiment it was found that the bias, if it was going to occur, occurred as the event description unfolded, it was not present immediately subjects had been given outcome knowledge before they read the event description.[3] If we are saying that hindsight bias results from the ease with which scenarios indicating the outcome which occurred can be imagined or brought to mind, as we are with the availability heuristic, this pattern of findings is consistent with this view. This is because subjects can only construct such scenarios as relevant and appropriate information is revealed to them. This means that outcome knowledge causes people to seek explanations as to why that outcome happened and encourages them to imagine ways in which it came about. The availability of these types of scenarios will increase the subjective likelihood of that outcome with hindsight bias as the result. Again, the availability heuristic is intimately bound up with a confirmatory hypothesis testing bias.

---

[3] If anchoring and adjustment were operating here we would expect people, on receipt of outcome knowledge before the event description, would be "anchored" to that outcome. This could mean that they see the outcome as 100% likely and adjust this assessment downwards in the light of conflicting information which is subsequently revealed. With this strategy hindsight bias should be present immediately, as it was not it seems reasonable to assume that such a strategy is not used.

The above discussion, by the same token, should also mean that hindsight bias should be present for outcomes which did not occur (i.e. those outcomes being seen as less likely in hindsight than foresight). As discussed in Section 2 of this Chapter, one of the most consistent findings of the experiments reported in this thesis was the absence of such an effect with such outcomes.

That there should be an effect here follows from the preceeding discussion because outcomes which did not occur should be more difficult to imagine happening as subjects should experience greater difficulty in constructing scenarios indicating their occurrence.[4] The failure to find such bias here, and in Fischhoff's research, provides additional substance to the methodological/design account given in the previous section.

In summary, the availability heuristic provides a good account of hindsight bias when it is shown through subjective likelihood assessments. It accounts for both the impact of outcome knowledge on judgements and the way in which information is made accessible from memory when making those judgements. We now turn to look at how the anchoring and adjustment heuristic explains hindsight bias when estimates are made.

---

4 In foresight the ease with which scenarios for each outcome can be imagined or constructed will only be influenced by the event description (plus any background knowledge the subject may have). In hindsight, by contrast, subjects will be influenced by outcome knowledge in addition to these factors.

14.2.2 Anchoring and Adjustment and Hindsight Bias

The anchoring and adjustment heuristic - where people make estimates by starting from an initial value that is adjusted to yield the final answer, is most appropriate to apply in situations where people make estimates rather than assessments.

Assessments are where people attempt to put some kind of formal figure on an internal state of knowledge about the world. This figure however, is not open to objective appraisal beyond consensus of opinion among a number of individuals. Estimates, by contrast, are where people put a formal figure on some property of the external world. This figure is open to objective appraisal.

Experiment 8 used factual material and, it will be recalled, provided the strongest evidence of all the experiments for hindsight bias. Providing hindsight subjects with knowledge of what the actual statistics were for the numbers of percentages of women in certain occupations/roles can be regarded as the anchor from which subjects adjust when making their own estimates. Indeed, with this task it is difficult to see what other kind of anchor, beyond the two limiting extremes, people could use. In foresight, of course, we would want to say, if anchoring and adjustment were operating, that one of these two extremes was used as the anchor from which adjustments were made to yield the final estimate.

Providing people with actual figures, which is difficult if not impossible to do with outcomes to social events, gives subjects such a natural starting point from which to base their estimates. The final estimates given by subjects, as shown by the results of Experiment 8, indicates that people make insufficient adjustment when attempting to ignore the fact that they know what the actual figures are.

Fischhoff (1975a), Fischhoff and Beyth (1975) both offered accounts of hindsight bias based on anchoring and adjustment. Neither of these experiments used material of a factual nature or asked subjects to make estimates. It is interesting to note that this early interpretation of the bias has been abandoned in subsequent research (e.g. Fischhoff, 1977) in favour of an interpretation using the availability heuristic.[5] However, what is often ignored is the fact that this heuristic may be the appropriate one to use in certain conditions. It is to this that we now turn.

14.2.3 Availability and Hindsight = Foresight

The availability heuristic is appropriately used in some circumstances, as a consequence, does not lead to bias. Consider the results of Experiment 5, the British Firemen's Strike, where it was found that asking subjects to generate their own outcomes and assess the likelihood of each, rather than assess the likelihood of outcomes specified by the experimenter, led to little or no bias.

Such results indicate that the "generate" strategy allows subjects in hindsight to more accurately reconstruct a foresight state. This implies that the availability of states of uncertainty experienced in foresight is greater when asked to express outcomes in their own words.

---

[5] Fischhoff (1977) uses factual material to investigate hindsight bias, as described in Chapter 3. However, subjects in these experiments were asked to make likelihood assessments about their chances of being right or wrong about some question of fact. They were not asked to estimate what they thought some statistic was, had they not known it. Hence the interpretation of hindsight bias is more appropriately given by availability than anchoring and adjustment.

This discussion, as with a previous one, relies upon their being a strong link between accessibility and availability. The assumption is plausible here for two reasons: (1) the generate strategy may entail subjects processing information, given to them at the time, more deeply than subjects in the "specify" treatment. As it has been shown that information processed in this way is remembered better on later occasions (Craik and Lockhart, 1972; Craik and Tulving, 1975) it is reasonable to assume that judgements made upon more deeply processed information will be better remembered or reconstructed. In the context of the availability heuristic, we can see that because this information is more accessible the judgement or state of uncertainty that existed in foresight wi..l be more "available" because it is easier to reconstruct or imagine. (2) Tulving and Thompson (1973); Tulving and Watkins (1975) have shown that if an initial judgement is remembered it may serve as a cuing mechanism and facilitate the recall of information that had been instrumental in reaching the judgement. In terms of foresight/ hindsight judgements this means that the availability of related information which resulted in a particular state of uncertainty will be more accessible if the judgement is remembered without the particular reasons for reaching this position in the first place.

These two arguments are really two sides of the same coin, one is saying that states of uncertainty experienced in foresight are made more available by the "generate" strategy because such a strategy makes people process the information needed to make the judgement more deeply. On the other side of the coin, the states of uncertainty experienced in foresight are more accurately reconstructed if the original judgement is remembered because it makes the information upon which the judgement was made more available from memory.

This discussion, as with a previous one, relies upon their being a strong link between accessibility and availability. The assumption is plausible here for two reasons: (1) the generate strategy may entail subjects processing information, given to them at the time, more deeply than subjects in the "specify" treatment. As it has been shown that information processed in this way is remembered better on later occasions (Craik and Lockhart, 1972; Craik and Tulving, 1975) it is reasonable to assume that judgements made upon more deeply processed information will be better remembered or reconstructed. In the context of the availability heuristic, we can see that because this information is more accessible the judgement or state of uncertainty that existed in foresight will be more "available" because it is easier to reconstruct or imagine. (2) Tulving and Thompson (1973); Tulving and Watkins (1975) have shown that if an initial judgement is remembered it may serve as a cuing mechanism and facilitate the recall of information that had been instrumental in reaching the judgement. In terms of foresight/hindsight judgements this means that the availability of related information which resulted in a particular state of uncertainty will be more accessible if the judgement is remembered without the particular reasons for reaching this position in the first place.

These two arguments are really two sides of the same coin, one is saying that states of uncertainty experienced in foresight are made more available by the "generate" strategy because such a strategy makes people process the information needed to make the judgement more deeply. On the other side of the coin, the states of uncertainty experienced in foresight are more accurately reconstructed if the original judgement is remembered because it makes the information upon which the judgement was made more available from memory.

In summary, the use of a "generate" type strategy encourages the appropriate use of the availability heuristic and hence results in little or no hindsight bias. This is because the strategy allows hindsight subjects to more accurately reconstruct states of uncertainty experienced in foresight.

The above discussion is specific to cases where subjects give both foresight and hindsight judgements. However, it is reasonable to assume that the first argument could apply just as well to cases where subjects are asked, in hindsight, to assess what they thought the likelihood of the outcomes were had they not known what had happened. This, of course, provided they are asked to generate their own outcomes.

We would predict, then, that subjects provided with an event description, told how it turned out, and asked to produce a number of possible outcomes (in their own words), including the one that happened, would show much less or no hindsight bias when compared with subjects who were provided with outcomes.

14.2.4 Interdependence of Availability and Representativeness

Much of the preceeding discussion has been concerned to demonstrate how the availability heuristic may be used to explain how outcome knowledge biases perception or not depending on conditions. One important aspect ignored so far is how the representativeness heuristic (the degree to which the outcome is similar to, or representative of, the features characteristic of the outcome) may play a role in determining the availability of instances or scenarios used when assessing likelihoods.

The ease with which instances or scenarios are available from

memory may, as Nisbett and Wilson (1978) point out, result from them being viewed as the most representative aspects of the event. This may be especially true when we consider outcome knowledge itself. Outcome knowledge, apart from being an important piece of information about an event, may also be regarded as the most representative piece of information about than event, if for no other reason that it being the outcome which happened. If this is true, then it is not hard to see that when trying to assess likelihoods the most easily imagined (available) scenarios which can be brought to mind will be those indicating the outcome.

This, in conjunction with Carroll's (1978) finding that people asked to imagine an event turning out in a certain way see that outcome as more likely than those not engaged on such a task and the confirmatory hypothesis - testing strategy that people appear to use (Snyder, in Press) provide a further explanation of hindsight bias.

In summary, not only may outcome knowledge influence what antecedent information is regarded as most representative of the event, but the outcome which occurred may itself be seen as the most representative piece of information. Both factors, due to representativeness, have important influences over the availability heuristic. The most available scenarios etc., being those which indicate the outcome which occurred, with the result that it is seen as more likely in hindsight than foresight.

14.2.5 General Considerations

Before considering implications for future research and decision-making, some general aspects of creeping determinism, hindsight bias, heuristics of thinking and the cognitive processes discussed in this thesis will be considered.

The creeping determinism/hindsight bias distinction, discussed and operationally defined in Chapter 2, is one which probably does not apply to historical or social events. In saying this, one is claiming that evidence of creeping determinism in such situations would, most of the time, be construed as hindsight bias. Justification for this comes from considering the cognitive task of someone provided with, or in possession of, outcome knowledge of an event and some information relating to the event before the outcome was known. Asking a person, with such knowledge, to assess the likelihood of that and alternative outcomes without specific instructions to ignore outcome knowledge may well cause the following sequence of thoughts: "what I am being asked to do is to judge the likelihood of the outcome which happened together with those that didn't. How can I best do this? I know what happened but it was not inevitable, or not 100% likely. However, knowing what happened may make it appear that way, so I shall ignore or try to ignore this, and try to make my likelihood assessments reflect only the information available before events turned out in such a way". Whilst this is highly stylised it does not seem to be an unreasonable speculation. It is backed up to a certain extent from introspective reports of subjects who were questioned by the author in the various creeping determinism experiments reported in this thesis.

The credibility of this together with the authors beliefs that the distinction between creeping determinism and hindsight bias is

no more than a technical one, except in special circumstances, means that any evidence of hindsight likelihood assessments for outcomes which occurred being greater than corresponding foresight assessments can be regarded as evidence of biased judgements. Such considerations presuppose the existence of such evidence, it is to this that we now turn.

The experiments reported in this thesis offer mixed support for hindsight judgements being biased. In view of this some discussion of the generality and strength of the bias is appropriate.

Generally, results from all the experiments investigating the effect of outcome knowledge on likelihood assessments are in the predicted direction (with only a few notable exceptions). Viewed from this perspective one might want to claim that hindsight bias is a general phenomenon. However, the effect does not appear to be a strong one. This latter point is the crucial one.

Time and again the results from the seven experiments fall into this kind of pattern. That is, results in the predicted direction but not, on the whole, significantly so. One might want to claim that the experiments did not take a large enough sample to warrant such a conclusion. Two answers can be offered here, firstly, large sample sizes, as used by Fischhoff and his colleagues, may produce significant results but only because smaller effects are found to be significant when sample size is increased (this point is dealt with at some length in Appendix 13). We are not only interested in significant results but also the size of the effect. Secondly, whilst smaller sample sizes reduce the power of a statistical test the consistency of the findings from the seven experiments reported in this thesis is enough to infer the effect to be a weak one.

Generally, then, one would conclude that hindsight bias is something that people are susceptible to but not to a great extent. The question then arises as to how important an influence on judgements and decision-making it is. On the whole it could be said that decision-making and prediction will not be overly influenced but knowledge of what went before in similar instances. However, it is reasonable to speculate that in special circumstances it may have an undesirably strong influence. In crisis decision-making in international politics for example.

In crisis decision-making the decision makers operate under a number of powerful constraints, perhaps the most important being time. A short time in which to make a crucial decision, means that a large amount of information has to be processed and assimilated. One simplication strategy which, according to May (1973) is extensively cued is to fit some past event to the present situation. Use of such analogies, which tend to oversimplify past and result in over-simplification of present events, may be where hindsight bias does play an important role. An undesirable outcome to the past event, which they generally are, may be seen as more inevitable than it was at the time. This results in the approach to the current crisis being perceived incorrectly with a strong possibility that the policy adopted here will be inappropriate.

Whilst hindsight bias may play an important role in such situations it is only because certain cognitive processes and/or heuristics of thinking are being misapplied also. Whilst, the interaction of hindsight bias and cognitive processes will not be considered further here, as some discussion is given in Chapter 2, this section will conclude by considering and speculating upon the generality and importance of the cognitive processes, particularly the heuristics of

thinking, detailed in Chapter 2 and earlier in this chapter.

Three issues deserve attention here: (1) how general is the application of a model of man as an intuitive scientist?; (2) how are the cognitive processes, heuristics of thinking to be regarded - are they conscious or unconscious processes?; (3) what relationship and interplay might we expect in the future between cognitive and social psychology? These issues will be treated in this order.

Firstly, the model of man as an intuitive scientist, using various heuristics, etc., to make judgements and decisions, to interpret and manage the information in the environment applies to everybody. One is not characterising the layman with operating with such processes and psychologists, scientists, etc., with other more rational or appropriate processes. Thus, for example, one can see evidence of the intuitive scientist in the professional scientist's activities. The work of Kuhn (1962); Lakatos (1970) and Feyerabend (1975) can be viewed in this way. Feyerabend in particular with his emphasis upon the importance of historical reconstruction could be analysed from a cognitive social psychological perspective. In a different way many of the criticisms of Freud's research procedures and his case summary method could also be reconceptualised using research from social cognition. Lastly, psychologists themselves are not to be regarded as immune.

Secondly, the heuristics of thinking and other cognitive processes, as Nisbett and Ross (1980), Nisbett and Wilson (1977) point out, are something we are not aware of using, and ones which most people would not admit to using. What status do such processes have then - are they unconscious, pre-conscious, non-conscious or what? Such issues have not been addressed by researchers to date and it is difficult to see

what kind of light could be thrown upon these cognitive processes by such a discussion. The best that can be said at present is that empirical evidence leads us to infer that people operate with such strategies but only because of the weight and breadth of the evidence. It is plausible that another model of man would be able to reconceptualise the findings. Such a development would be akin to a Kuhnian paradigm shift i.e. there would be no logical grounds for researchers in the old paradigm to accept the new one. Still this seems to be the way science progresses.

Thirdly, and lastly, the most exciting aspect of this work, as a psychologist, is the development of strong links between cognitive and social psychology. Such an interaction between the two areas is beneficial to both. For cognitive psychology an extended quote from Neisser (1976)[1] demonstrates this clearly. He says:-

> "The study of information processing has momentum and prestige, but it has not yet committed itself to any conception of human nature that could apply beyond the confines of the laboratory. And within that laboratory, its basic assumptions go further than the computer model to which it owes its existence. There is still no account of how people act in or interact with the ordinary world. Indeed, the assumptions that underlie most contemporary work on information processing are surprisingly like those of nineteenth-century introspective psychology, though without introspection itself.

---

[1] U. Neisser (1976) Cognition and Reality, San Francisco: Freeman and Co.

what kind of light could be thrown upon these cognitive processes by such a discussion. The best that can be said at present is that empirical evidence leads us to infer that people operate with such strategies but only because of the weight and breadth of the evidence. It is plausible that another model of man would be able to reconceptualise the findings. Such a development would be akin to a Kuhnian paradigm shift i.e. there would be no logical grounds for researchers in the old paradigm to accept the new one. Still this seems to be the way science progresses.

Thirdly, and lastly, the most exciting aspect of this work, as a psychologist, is the development of strong links between cognitive and social psychology. Such an interaction between the two areas is beneficial to both. For cognitive psychology an extended quote from Neisser (1976)[1] demonstrates this clearly. He says:-

> "The study of information processing has momentum and prestige, but it has not yet committed itself to any conception of human nature that could apply beyond the confines of the laboratory. And within that laboratory, its basic assumptions go further than the computer model to which it owes its existence. There is still no account of how people act in or interact with the ordinary world. Indeed, the assumptions that underlie most contemporary work on information processing are surprisingly like those of nineteenth-century introspective psychology, though without introspection itself.

---

[1] U. Neisser (1976) Cognition and Reality, San Francisco: Freeman and Co.

> If cognitive psychology commits itself too thoroughly to this model, there may be trouble ahead. Lacking in ecological validity, indifferent to culture, even missing some of the main features of perception and memory as they occur in ordinary life, such a psychology could become a narrow and uninteresting specialised field".
>
> (pp. 6-7).

A strong partnership between cognitive and social psychology would prevent cognitive psychology ending up in obscurity. Social psychology can offer relevance and, perhaps more importantly, behavioural consequences of cognitive psychology. Social psychology can also provide ecological validity by being a regulatory force.

How does social psychology benefit from cognitive psychology? Primarily it offers a depth of analysis and understanding which, until recently has been lacking in social psychologists. Social psychologists are now in a position to offer cognitive explanations, as well as environmental, affective, etc., for social behaviour.

Finally, this relationship between cognitive and social psychology, because it is a new one, has much more to achieve than that stated above. Many traditional areas of social psychology could benefit from research by cognitive psychologists with their specialised techniques. For example, the notion of cognitive consistency is an important and persuasive one in social psychology, yet it seems strange that such a cognitive principle has not received substantial attention and investigation by cognitive psychologists.

In conclusion, a more productive psychology of social cognition can be produced by cognitive psychologists receiving inputs from social psychology and social psychology receiving inputs from cognitive psychology. A closer liaison between the two disciplines will, I believe, result in a more worthwhile and exciting discipline of psychology.

14.2.6 Summary

It has been argued here that hindsight bias, when it occurs, may be best explained by the inappropriate use of the availability heuristic in conjunction with the representativeness heuristic. This is so when subjects are asked to make likelihood assessments. In cases where subjects show hindsight bias when estimating actual figures anchoring and adjustment provides the most plausible explanation.

The availability heuristic was also shown to be appropriately used under certain conditions with the result that little or no hindsight bias is detected.

The experiments reported in this thesis, then, have not only provided information concerning the conditions under which hindsight bias is or is not likely to appear, but have also provided some indication of under what experimental conditions heuristics of thinking produce biased judgements and the conditions under which it might be appropriate to use these heuristics.

14.3 IMPLICATIONS FOR FUTURE RESEARCH

One area which has not been empirically investigated in this thesis, but has been hinted at in several of the experimental reports, is to do with the way in which information search may be guided by outcome knowledge.

If, as has been shown, hindsight bias is affected by the type and order of information we would expect that a person who knows how an event turned out, but does not have sufficient knowledge about the event to assess how likely it was, would show distinct patterns

of information search. We might hypothesise that information indicating the outcome which occurred, and not that indicating other outcomes, will be more actively searched for. Two reasons would argue for this; (1) Fischhoff's (1975a) findings that subjects perceive the most relevant information to be that confirming the outcome which occurred. (2) A confirmatory hypothesis-testing strategy which people use, already discussed in this chapter.

The way in which people search for information when making some judgement about the past or, perhaps more importantly, when attempting to predict future events based on one's knowledge of the past is of obvious practical importance for decision-making at all levels in society. Furthermore, in view of the overwhelming amount of information easily available to us (and potentially more easily available because of the microprocessor revolution) it will become essential to discover heuristic strategies used in search and the potential bias they may be subject to. With this knowledge, psychologists may be able to offer prescriptions as to ways in which information may be searched for which more correctly characterises and simplifies the world.

Not only may decisions, judgements and perceptions be influenced by outcome knowledge and the search for information but they may also be affected by the order in which information is searched for.

The experiments reported in this thesis offer only mixed support for order of information affecting judgements (both in relation to outcome knowledge and on their own) future research might build upon these findings. For example, the order in which different types of

information, such as those in Experiment 5, are searched for may be an important factor. We might, speculatively, say that searching for information surrounding the outcome first and for early events second may bias judgements more than if the other way around.

Finally, further research is needed into the conditions under which hindsight bias appears and does not appear, especially in "applied" situations. For example, does the amount of information describing a scenario affect hindsight bias? The experiments reported here presented only short (about 150 word) summaries to subjects. Whilst this may adequately simulate some judgements made in "real-life", there are obviously many important instances in which judgements are based on much larger amounts of information. Before we could generalise the findings reported here to such conditions further research into this is required.

14.4 CONCLUDING REMARKS

Although it is not the purpose of this thesis to offer prescriptions we might conclude by briefly considering some practical implications of the research reported here.

Decision makers who use historical knowledge in predicting the future or people who attempt to understand the past need, generally, to guard against hindsight bias. On the basis of the empirical findings reported here we might suggest that such people try to see and explain how events could have turned out in ways other than they did. Also, we would suggest that strategies which encourage a deeper level of information-processing should be encouraged.

Furthermore, if people are in possession of information surrounding an event and outcome knowledge, but have not had occasion to read either, they should be encouraged to read the antecedent details first, formulate their own predictions and likelihood assessments, perhaps writing them down, before reading what the outcome was. By this argument, experimental reports in psychology, or elsewhere for that matter, should have the abstract last and not first.

On a different note, where commitment or desire to see certain outcomes happen is high people may be subject to a bias as severe to that of hindsight. In such situations people should make every effort to become aware of their own biases and reach beyond what they would personally like to see happen.

14.5 FINAL NOTE

Finally, and to conclude on a general note, much of the future research in cognitive social psychology, as characterised in Chapter 2, will be devoting its attention, like this thesis, to establishing conditions under which biases and errors occur together with when it is appropriate to use certain heuristics of thinking and when it is not.

Such research, because of its potential value to decision makers, should not only be reported for consumption by psychologists but should be made relevant to people for whom it has practical significance.

REFERENCES

ABELSON, R.P. (1976). Script processing in attitude formation and decision making. In J.S. Carroll & J.W. Payne, Cognition and Social Behaviour. Hillsdale, New Jersey : Lawrence Erlbaum Associates.

ABELSON, R.P. (1974). Social psychology's rational man. In G.W. Mortimer & S.I. Benns (Eds.), The Concept of Rationality in the Social Sciences. London : Routledge and Kegan Paul.

ABELSON, R.P. & ROSENBERG, M.J. (1958). Symbolic psycho-logic : A model of altitudinal cognition. Behavioural Science. 3, 1-13.

AJZEN, I. (1977). Intuitive theories of events and the effects of bare-rate information on prediction. Journal of Personality and Social Psychology, 35, 303-314.

ANDERSON, N.H. (1968). Application of a linear-serial model to a personality impression task using serial presentation. Journal of Personality and Social Psychology. 10, 354-362.

ANDERSON, N.H. & BARRIOS, A.A. (1961). Primacy effects in personality and impression formation. Journal of Abnormal and Social Psychology, 63, 346-350.

ANDERSON, N.H. & HUBERT, S. (1963). Effects of concomitant verbal recall on order effects in personality impression formation. Journal of Verbal Learning and Verbal Behaviour. 2, 379-391.

ANDERSON, N.H. & JACOBSON, A. (1965). Effect of stimulus inconsistency and discounting instructions in personality impression formation. Journal of Personality and Social Psychology. 2, 531-539

ANDERSON, N.H. & NORMAN, A. (1964). Order effects in impression formation in four classes of stimuli. Journal of Abnormal and Social Psychology. 69, 467-471.

ASCH, S.E. (1946). Forming impressions of personality. Journal of Abnormal and Social Psychology. 41, 258-290.

AXELROD, R. (1976)(Ed.). Structure of Decision : Cognitive Maps of Political Elites. Princetown : Princetown University Press.

BACON, F. (1620). The New Organon and Related Writings. New York : Liberal Arts Press, 1960.

BAR-HILLEL, M. (in Press). What features make samples seem representative? Journal of Experimental Psychology : Human Perception and Performance.

BARNETT, C. (1963). The Sword Bearers : Studies in Supreme Command in the First World War. London : Eyre & Spottiswoode.

BECKMAN, L. (1970). Effects of students' performance on teachers' and observers' atributions. Journal of Educational Psychology. 65, 198-204

BERMANT, G., MCGUIRE, M., MCKINLEY, W. & SALO, C. (1974) The logic of simulation in jury research. Criminal Justice and Behaviour 1, 224-233

BERMANT, G., NEMETH, C. & VIDMAR, N. (1976). Psychology and the law : Status and challenge. In G. Bermant, C. Nemeth & N. Vidmar (Eds.). Psychology and the Law. Lexington Books.

BIERBRAUER, G. (1979). Why did he do it? Attribution of obedience and the phenomena of dispositional bias. European Journal of Social Psychology, 9, 67-84.

BORGIDA, E. & NISBETT, R.E. (1977). The differential impact of abstract versus concrete information on decisions. Journal of Applied Social Psychology, 7, 258-271

BRADLEY, G.W. (1978). Self-serving biases in the attribution process : A re-examination of the fact or fiction question. Journal of Personality and Social Psychology, 36, 56-71.

BREHM, J.W. (1956). Post-decisional changes in the desirability of alternatives. Journal of Abnormal and Social Psychology, 52, 384-389.

BRUNER, J. (1957). On perceptual readiness. Psychological Review, 64, 129-130.

BRUNSWICK, E. (1947). Systematic and Representative Design of Psychological Experiments in Physical and Social Perception. Berkeley : University of California Press.

BUSCH, F. (1961). Trial Procedure Materials Indianapolis : Bobbs-Merrill.

BUSS, A.R. (1978). Reasons and causes in attribution theory : A conceptual critique. Journal of Personality and Social Psychology, 36, 1311-1321.

BUTLER, D. & KAVANAGH, D. (1975). The British General Election of October 1974. London : The Macmillan Press Ltd.

CABRERA, H.A. (1969). A comprehensive evaluation of pregnancy tests. American Journal of Obstetrics and Gynaecology, 32-38.

CANTRIL, H. (1938). The prediction of social events. Journal of Abnormal and Social Psychology, 33, 364-389.

CARROLL, J.S. (1978). The effect of imagining an event on expectations for the event : An interpretation in terms of the availability heuristic. Journal of Experimental Social Psychology, 14, 88-96

CARROLL, J.S. & SIEGLER, R.S. (1977). Strategies for the use of base-rate information. Organisational Behaviour and Human Performance, 19, 392-402.

CHAPMAN, L.J. & CHAPMAN, J.P. (1969) Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. Journal of Abnormal Psychology, 74, 271-280

CHAPMAN, L.J. & CHAPMAN, J.P. (1967). Genesis of popular but erroneous diagnostic observations. Journal of Abnormal Psychology, 72, 193-204.

COHEN, J. (1969). Statistical Power Analysis in the Behavioural Sciences. New York : Academic Press.

COHEN, J. (1962). The statistical power of abnormal-social psychological research. Journal of Abnormal and Social Psychology, 65, 145-153.

COHEN, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37-46.

COHEN, L.J. (1979a). On the psychology of prediction : Whose is the fallacy? Cognition, 7, 385-407

COHEN, L.J. (1979b). Whose is the fallacy? : a rejoinder to Daniel Kahneman and Amos Tversky. Cognition, 8, 89-92.

CONLEY, J.M., BARR, W.M. & LIND, E.A. (1978). The power of language : Presentational style in the courtroom. Duke Law Journal, 6, 1375-1398.

COOMBS, B. & SLOVIC, P. (1978). Causes of death : Biased newspaper coverage and biased judgements. Decision Research Report, 8

COSTOPOULOUS, W.C. (1972). Persuasion in the courtroom. Duquesne Law Review, 10, 384-409.

CRAIK, F.I.M. & LOCKHART, R.S. (1972). Levels of processing : A framework for memory search. Journal of Verbal Learning and Verbal Behaviour, 11, 671-684.

CRAIK, F.I.M. & TULVING, E. (1975). Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology : General, 104, 268-294.

CROZIER, W.R. (1979). The interaction of value and subjective probability in risky decision-making. British Journal of Psychology 70, 489-495.

DALE, H.C.A. (1968). Weighing evidence : An attempt to assess the efficiency of the human operator. Ergonomics, 11, 215-230.

DAVIS, J.H. (1976). Group decision and procedural justice. Paper presented at Otterbeuren Symposium on Group Decision Making. Otterbeuren, Federal Republic of Germany. September.

DAWES, R. (1976). Shallow Psychology. In J. Carroll & J. Payne (Eds.) Cognition and Social Behaviour. Hillsdale, N.J. : Erlbaum.

DEVLIN, P. (1966). Trial by Jury. London : Stevens & Sons Ltd.

EHRLICH, D., GUTTMAN, I., SCHONBACH, P. & MILLS, J. (1957). Post-decision exposure to relevant information. Journal of Abnormal and Social Psychology, 54, 98-102.

EINHORN, H.J. & HOGARTH, R.M. (1978). Confidence in judgement : Persistence of the illusion of validity. Psychological Review, 85, 395-416.

EISER, J.R. (1980). Social risk taking, conservatism and the social trap. Paper presented to a European Association of Experimental Social Psychology - International Workshop on Social Psychology and Social Policy. University of Kent at Canterbury, 29th March - 3rd April.

ELTON, G.R. (1967). The Practice of History. London : Methuen & Co.

ERICKSON, B., LIND, E.A., JOHNSON, B.C. & BARR, W.M. (1978). Speech style and impression formation in a court setting: The effects of "powerful" amd "powerless" speech. Journal of Experimental Social Psychology, 14, 266-279.

FELLNER, C.H. & MARSHALL, J.R. (1970). Kidney Donors. In J. McCauley & L. Berkowitz (Eds.), Altruism and Helping Behaviour. New York : Academic Press.

FESTINGER, L. (1957). A Theory of Cognitive Dissonance. Stanford, California : Stanford University Press.

FESTINGER, L. (1964) (Ed.). Conflict, Decision and Dissonance. Stanford, California : Stanford University Press.

FEYERABEND, P. (1975). Against Method : Outline of an Anarchistic Theory of Knowledge. London : New Left Books.

FISCHHOFF, B. (1980). For those condemned to study the past : Reflections on historical judgement. In R.A. Sheveder & D.W. Fiske, (Eds.). New Directions for Methodology of Behavioural Science : Judgement in Behavioural Research. San Francisco : Jossey-Bass.

FISCHHOFF, B. (1977). Perceived informativeness of facts. Journal of Experimental Psychology : Human Perception and Performance. 3, 349-358.

FISCHHOFF, B. (1976). Attribution theory and judgement under uncertainty. In J.H. Harvey, W.J. Ickes & R.F. Kidd (Eds.), New Directions in Attribution Research. New Jersey : Lawrence Erlbaum Associates.

FISCHHOFF, B. (1975b) The silly certainty of hindsight. Psychology Today. April, 71-76

FISCHHOFF. B. (1975a). Hindsight≠foresight : The effect of outcome knowledge on judgement under uncertainty. Journal of Experimental Psychology : Human Perception and Performance, 1, 288-299.

FISCHHOFF, B. & BEYTH, M. (1975). "I knew it would happen" : Remembered probabilities of once future things. Organisational Behaviour and Human Performance. 13, 1-16.

FISCHHOFF, B. & MAROM, B. (1976). Failure has many fathers. Policy Sciences. 6. 375-384.

FISCHHOFF, B. & SLOVIC, P. (1978). A little learning ... : Confidence in multicue judgements. Attention and Performance VIII : Princetown, New Jersey.

FISCHHOFF, B. & SLOVIC, P. (1976). The certainty illusion. Oregan Research Bulletin. Vol. 16, No. 4.

FISCHHOFF, B., SLOVIC, P. & LICHTENSTEIN, S. (1979). Subjective sensitivity analysis. Organisational Behaviour and Human Performance, 23, 339-359.

FISCHHOFF, B., SLOVIC, P. & LICHTENSTEIN, S.(1977). Knowing with certainty : The appropriateness of extreme confidence. Journal of Experimental Psychology : Human Perception and Performance. 3, 552-564.

FISHBEIN, M. & AJZEN, I. (1975). Belief, Attitude, Intention and Behaviour : An Introduction to Theory and Research. London : Addison Wesley.

FOSS, R.D. (1975). A critique of jury simulation research. Paper presented at the annual meeting of the American Psychological Association, Chicago, August 30th - September 3rd.

FRIEZE, I. & WEINER, B. (1971). Cue utilization and attributional judgements for success and failure. Journal of Personality. 39, 591-605.

GOETHALS, G.R., ALLISON, S.J. & FROST, M. (1979). Perceptions of the magnitude and diversity of social support. Journal of Experimental Social Psychology. 15, 570-581.

GOETHALS, G.R. & DARLEY, J.M. (1977). Social Comparision Theory : An attributional approach. In J.M. Suls & R.L. Miller (Eds.), Social Comparision Processes : Theoretical and Empirical Perspectives. Washington D.C. : Hemmisphere.

GOFFMAN, E. (1963). Stigma : Notes on the Management of Spoiled Identity. Englewood Cliffs : Prentice Hall.

GULLEY, H.E. & BERLO, D.K. (1956). Effect of intercellular and intra cellular speech structure on attitude change and learning. Speech Monographs. 23, 288-297.

HABER, R. (1966). Nature of the effect of set perception. Psychological Review, 73, 335-351.

HANSEN, R.D. & DONOGHUE, J.M. (1977). The power of consensus : information derived from one's own and others' behaviour. Journal of Personality and Social Psychology. 35, 294-302.

HAYS, W.L. (1973). Statistics for the Social Sciences. Second Edition. London : Holt, Rinehart and Winston.

HEIDER, F. (1958). The Psychology of Interpersonal Relations. New York : Wiley.

HEIDER, F. (1946). Attitudes and cognitive organisation. Journal of Psychology. 21, 107-112.

HENDRICK, C. & CONSTANTINI, A.F. (1970). Effects of varying trait inconsistency and responce requirements on the primacy effect in impression formation. Journal of Personality and Social Psychology. 15, 158-167.

HERMANN, C.F. (1972)(Ed.). International Crises : Insights from Behavioural Research. London : The Free Press.

HOCHBERG, B.C. & STIRES, L.K. (1973). The effect of several types of pre-trial publicity on the guilt attributions of simulated jurors. Journal of Applied Social Psychology. 3, 267-275.

HOVLAND, C.I. (1957) (Ed.). The Order of Presentation in Persuasion. New Haven : Yale University Press.

HOVLAND, C.I., JANIS, I.L. & KELLEY, H.H. (1953). Communication and Persuasion : Psychological Studies of Opinion Change. New Haven : Yale University Press.

HOVLAND, C.I. & MANDELL, W. (1952). An experimental comparsion of conclusion drawing by the communicator and by the audience. Journal of Abnormal Social Psychology. 47, 581-588.

JANIS, I. (1972). Victims of Groupthink : A Psychological Study of Foreign Policy Decisions and Fiascos. Boston : Houghton Mifflin Co.

JANIS, I.L. & HOVLAND, C.I. (1959). An overview of persuasibility research. In C.I. Hovland & I.L. Jones (Eds.), Personality and Persuasibility. New Haven : Yale University Press.

JANIS, I.L. & HOVLAND, C.I. (1959) (Eds.), Personality and Persuasibility. New Haven : Yale University Press.

JANIS, I.L. & MANN, L. (1977). Decision Making : A Psychological Analysis of Conflict, Choice and Commitment. New York : Free Press.

JERVIS, R. (1976). Perception and Misperception in International Politics. Princetown : Princetown University Press.

JONES, E.E. (1976). How do people perceive the causes of behaviour? American Scientist, 64, 300-305

JONES, E.E. & DAVIS, K.E. (1965). From acts to dispositions : The attribution process in person perception. In L. Berkowitz (Ed.) Advances in Experimental Social Psychology, Vol. 2. New York : Academic Press.

JONES, E.E. & GOETHALS, G.R. (1972). Order effects in impression formation : Attribution context and the nature of the entity. In E.E. Jones et al (Eds.), Attribution : Perceiving the Causes of Behaviour. Morristown, N.J. : General Learning Press.

JONES, E.E., GOETHALS, G.R., KENNINGTON, G.E. & SEVERENCE, L.J. (1972). Primacy and assimilation in the attribution process : The stable entity proposition. Journal of Personality. 40, 250-274.

JONES, E.E. & HARRIS, V.A. (1967). The attribution of attitudes. Journal of Experimental Social Psychology. 3, 1-24

JONES, E.E. & KOHLER, R. (1958). The effects of plausibility on the learning of controversial statements. Journal of Abnormal and Social Psychology, 57, 315-320.

JONES, E.E. & NISBETT, R.E. (1971). The actor and the observer : Divergent perceptions of the causes of behaviour. In E.E. Jones, D.E. Kanouse, H.H. Kelley, R.E. Nisbett, S. Valins & B. Weiner, (Eds.), Attribution : Perceiving the causes of behaviour. Morristown, N.J. : General Learning Press.

JONES, E.E, ROCK,L., SHAVER, K.G., GOETHALS, G.R. & WARD, L.M. (1968). Pattern of performance and ability attribution : An unexpected primacy effect. Journal of Personality and Social Psychology, 10, 317-341.

JONES, R.A. (1977). Self-Fulfilling Prophecies : Social Psychological and Physiological Effects of Expectancies. Hillsdale, New Jersey : Lawrence Erlbaum Associates.

KAHNEMAN, D. & TVERSKY, A. (in Press). Prospect theory : An analysis of decision under risk. Econometrica.

KAHNEMAN, D. & TVERSKY, A. (1979). Intuitive prediction : Biases and corrective procedures. Management Science. in Press

KAHNEMAN, D. & TVERSKY, A. (1973). On the psychology of prediction. Psychological Review, 80, 237-251.

KAHNEMAN, D. & TVERSKY, A. (1972). Subjective probability : A judgement of representativeness. Cognitive Psychology. 3, 430-454.

KELLEY, H.H. (1972). Attribution in social interaction. In E.E. Jones, D.E. Kanouse, H.H. Kelley, R.E. Nisbett, S. Valins & B. Weiner, (Eds.). Attribution : Perceiving the Causes of Behaviour. New Jersey : General Learning Press.

KELLEY, H.H. (1967). Attribution theory in social psychology. In Nebrask Symposium on Motivation. Lincoln : University of Nebraska Press. pp 192-238.

KELMAN, H.C. (1967). Human use of human subjects : The problem of deception in social psychological experiments. Psychological Bulletin. 67, 1-11.

KEPPEL, G. (1973). Design and Analysis : A Researcher's Handbook New Jersey : Prentice Hall.

KERBER, I., INCLAN, A., FOWLER, E., DAVIS, I.T. & FISH, S. (1970). Immunological tests for pregnancy. Obstetrics and Gynaecology. 36, 37-43.

KERR, N.L., ATKINS, R.S., STASSER, G., MEEK, D., HOLT, R.W. & DAVIS, J.H. (1976). Guilt beyond a reasonable doubt : Effects of concept definition and assigned decision rule on the judgement of mock jurors. Journal of Personality and Social Psychology. 34, 282-294.

KINDER, D.R. & WEISS, J.A. (1979). In lieu of rationality : Psychological perspectives on foreign policy decision-making. Journal of Conflict Resolution, 22, 707-735.

KORIAT, A. (in Press). Reasons for confidence. Journal of Experiment Psychology : Human Learning and Memory.

KRUGLANSKI, A.W. (1970). Attributing trustworthiness in supervisor-worker relations. Journal of Experimental Social Psychology, 6, 214-232.

KUHN, T.S. (1962). The Structure of Scientific Revolutions. Chicago : University of Chicago Press.

LAKATOS, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.). Criticism and the Growth of Knowledge. Cambridge University Press.

LANA, R.E. (1963b). Interest, Media and order effects in persuasive communications. Journal of Psychology, 56, 9-13.

LANA, R.E. (1963a). Controversy of the topic and the order of presentation in persuasive communications. Psychological Reports, 12, 163-170.

LANA, R.E. (1961). Familiarity and the order of presentation of persuasive communications. Journal of Abnormal and Social Psychology, 162, 573-577.

LAWSON, R.G. (1970). Experimental research on the organisation of persuasive arguments : An application to courtroom communications. Law and Social Order 579-608.

LAWSON, R.G. (1969). The law of primacy in the criminal courtroom. TheJournal of Social Psychology, 77, 121-131.

LAWSON, R.G. (1967). Order of presentation as a factor in jury persuasion. Kentucky Law Journal, Vol. 56, 523-555.

LICHTENSTEIN, S. & FISCHHOFF, B. (1977). Do those who know more also know more about how much they know? Organisational Behaviour and Human Performance, 20, 159-183.

LICHTENSTEIN, S., FISCHHOFF, B & PHILLIPS, L. (1977). Calibration of probabilities : The state of the art. In Proceedings of the Fifth Research Conference on Subjective Probability, Utility and Decision Making. H. Jungermann & G. de Zeeuw (Eds.)

LICHTENSTEIN, S., SLOVIC, P., FISCHHOFF, B., LAYMAN, M. & COOMBS, B. (1978). Judged frequency of lethal events. Journal of Experimental Psychology : Human Learning and Memory, 4, 551-578.

LINTON, M. & GALLO, P.S. (1975). The Practical Statistician : Simplified Handbook of Statistics. Monterey, California : Brooks/Cole Publishing Company.

LOCKE, D. & PENNINGTON, D.C. (1980). Shop-soiled goods : Reasons and causes in attribution theory. Unpublished research report.

LORD, G., ROSS, L. & Lepper, M.R. (1979). Biased assimilation and attitude polarisation : The effects of prior theories on subsequently considered evidence. Journal of Personality and Social Psychology, 37, 2098-2109.

LUCHINS, A.S. (1965). Anchorage and order effects of information in personality impression. Journal of Social Psychology, 66, 1-14.

LUCHINS, A.S. (1962). Primacy - recency in communications reflecting attitudes towards segregation. Journal of Social Psychology, 58, 357-369.

LUCHINS, A.S. (1957). Primacy-recency in impression formation. In C.I. Hovland et al (Eds.). The Order of Presentation in Persuasion, New Haven, Conn. : Yale University Press.

LUCHINS, A.S. & LUCHINS, E.H. (1970). Effects of preconceptions and communications on impressions of a person. Journal of Social Psychology, 81, 243-252.

LUND, F. (1925). The psychology of belief : A study of its emotional and volitional determinants. Journal of Abnormal and Social Psychology, 20, 174-196.

MAHONEY, M.J. (1977). Publication prejudices : An experimental study of confirmatory bias in the peer review system. Cognitive Therapy and Research, 1, 161-175.

MAIER, N.R.F. & MAIER, R.A. (1957). An experimental test of the effects of 'developmental' versus 'free' discussion on the quality of group decisions. Journal of Applied Psychology, 41, 320-323.

MANIS, M., DOVALINA, I., AVIS, N.E. & CARDOZE, S. (1980). Base rates can affect individual predictions. Journal of Personality and Social Psychology, 38, 231-248.

MANIS, M. & PLATT, M.B. (1979). Order effects in the integration of verbal descriptions. Personality and Social Psychology Bulletin, 5, 57-60.

MASSAD, C.M., HUBBARD, M. & NEWTSON, D. (1979). Selective perception of events. Journal of Experimental Social Psychology, 15, 513-532.

MAY, E.R. (1973). "Lessons" of the past : The Use and Misuse of History in American Foreign Policy. London : Oxford University Press.

MAYO, R.W. & THOMPSON, R. (1965). Comparison of pregancy tests. Obstetrics and Gynaecology, 25, 699-704.

MCARTHUR, L.A. (1972). The how and what of why. Journal of Personality and Social Psychology, 22, 171-193.

MCCAULEY, C. & JAQUES, S. (1979). The popularity of conspiracy theories of presidential assassination. Journal of Personality and Social Psychology, 37, 637-644.

MCCORDUCK, P. (1979). Machines Who Think. San Francisco : W.H. Freeman and Company.

MCGREGOR, D. (1938). The major determinants of the prediction of social events. Journal of Abnormal and Social Psychology, 33, 179-204

MCGUIRE, W.J. (1969). The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.). The Handbook of Social Psychology, Vol. 3. Reading, Massachusetts : Addison-Wesley Publishing Co.

MCKELVY, W.R. (1952). Presentation of the case in chief. In J. Appleman (Ed.), Successful Jury Trials. Indianapolis : Bobs Merrill.

MCKIE, D., COOK, C. & PHILLIPS, M. (1978). The Guardian Quartet Election Guide. London : Quartet Books.

MILBURN, M.A. (1978). Sources of bias in the prediction of future events. Organisational Behaviour and Human Performance, 21, 17-26.

MILGRAM, S. (1963). Behavioural study of Obediance. Journal of Abnormal and Social Psychology. 67, 371-378.

MILLER, A.J. (1968). Inference and prediction of events and event generators. Journal of Experimental Psychology, 76, 147-153.

MILLER, D.T. (1976). Ego involvement and attributions for success and failure. Journal of Personality and Social Psychology, 34, 901-906.

MILLER, D.T. & ROSS, M. (1975). Self-serving biases in the attribution of causality : Fact or fiction? Psychological Bulletin, 82, 213-225.

MILLER, G.A. (1956). The magic number seven plus or minus two : some limits on our capacity to process information. Psychological Review, 63, 81-97.

MISCHEL, W. (1979). On the interface of cognition and personality : Beyond the person situation debate. American Psychologist, 34, 740-754

MISCHEL, W. (1968). Personality and Assessment. New York : Wiley.

MORLEY, I.E. (1980). Preparation for negotiation : Conflict, choice and commitment. Unpublished research report, University of Warwick.

NEMETH, C. (1977). Interaction between jurors as a function of majority vs unanimity decision rules. Journal of Applied Social Psychology. 7, 38-56.

NEMETH, C. & SOSIS, R.H. (1973). A simulated jury study : characteristics of the defendant and the jurors. Journal of Social Psychology, 90, 221-229.

NEWCOMBE, T.M. (1959). Individual systems of orientation. In Psychology : A study of Science, Vol. 3. Edited by E. Koch. New York : McGraw-Hill.

NISBETT, R.E. & BORGIDA, E. (1975). Attribution and the psychology of prediction. Journal of Personality and Social Psychology, 32, 932-943.

NISBETT, R.E., BORGIDA, E., CRANDELL, R. & REED, M. (1976). Popular induction : Information is not always informative. In J.S. Carroll & J.W. Payne (Eds.), Cognition and Social Behaviour. Hillsdale, New Jersey : Erlbaum.

NISBETT, R.E., CAPUTO, C., LEGANT, P. & MARACEK, J. (1973). Behaviour as seen by the actor and as seen by the observer. Journal of Personality and Social Psychology, 27, 154-164.

NISBETT, R. & ROSS, L. (1980). Human Inference : Strategies and Shortcomings of Social Judgement. New Jersey : Prentice Hall.

NISBETT, R.E. & WILSON, T.D. (1977). Telling more than we can know : Verbal reports on mental processes. Psychological Review, 84. 231-259

NOGOUERES, H. (1963). Munich : The Phoney Peace. London : Weidenfeld & Nicolson.

OSGOOD, E.C., SOCI, G.J. & TANNENBAUM, P.H. (1957). The Measurement of Meaning. Urbana : University of Illinois Press.

OSGOOD, E.C. & TANNENBAUM, P.H. (1955). The principle of congruity in the prediction of attitude change. Psychological Review, 65, 42-55.

ORNE, M.T. (1969). Demand characteristics and the concept of quasi-controls. In R. Rosenthal & R.L. Rosnow (Eds.). Artifact in Behavioural Research. New York : Academic Press.

ORNE, M.T. (1962). On the social psychology of the psychological experiment : With particular reference to demand characteristics and their implications. American Psychologist, 17, 776-783.

ORNE, M.T. (1959). The demand characteristics of an experimental design and their implications. Paper presented at the meeting of the Americal Psychological Association, Cincinnati.

PENROD, S. & HASTIE, R. (1979). Models of jury decision making : A critical review. Psychological Bulletin, 86, 462-492.

PETERSON, C.R. & BEACH, L.R. (1967). Man as an intuitive statistician. Psychological Bulletin, 68, 29-46.

PETERSON, C.R. & DUCHARME, W.M. (1967). A primacy effect in subjective probability revision. Journal of Experimental Psychology. 73, 61-65.

PETERSON, C,R., SCHNEIDER, R.J. & MILLER, A.J. (1965). Sample size and the revision of subjective probabilities. Journal of Experimental Psychology, 69, 522-527.

PFEFFER, J. (1978). Organisational Design. Arlington Heights. Illinois : A.H.M. Publishing Corporation.

PHILLIPS, L.D. & EDWARDS, W. (1966). Conservatism in a simple probability inference task. Journal of Experimental Psychology. 72, 346-357.

PIAGET, J. (1929). The Child's Conception of the World. London : Routledge and Kegan Paul.

PITZ, G.F. & REINHOLD, H. (1968). Pay-off effects in sequential decision -making. Journal of Experimental Psychology, 77, 249-257.

PRUITT, D.G. & HOGE, R.D. (1965). Strength of the relationship between an event and its subjective probability as a function of the method of measurement. Journal of Experimental Psychology, 69, 483-489.

ROBY, T.B. (1967). Belief states and sequential evidence. Journal of Experimental Psychology, 75, 236-245.

ROSENTHAL, R. (1967). Covert communication in the psychology experiment. Psychological Bulletin, 67, 356-367.

ROSENTHAL, R. (1966). Experimenter Effects in Behavioural Research. New York : Appleton-Century-Crofts.

ROSENTHAL, R. (1965). The volunteer subject. Human Relations, 18, 389-406.

ROSENTHAL, R. & GAITO, J. (1963). The interpretation of levels of significance by psychological researchers. Journal of Psychology, 55, 33-38.

ROSNOW, R.L. & GOLDSTEIN, J. (1967). Familiarity, salience and the order of presentation of communications. Journal of Social Psychology, 73, 97-110.

ROSNOW, R.L., HOLZ, R. & LEVIN, J. (1966). Differential effects of complimentery and competing variables in primacy-recency. Journal of Social Psychology, 69, 135-147.

ROSS, L. (1977a) The intuitive psychologist and his shortcomings : Distortions in the attribution process. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology, Vol. 10, 173-220.

ROSS, L. (1977b). Some afterthoughts on the intuitive psychologist. In L. Berkowitz, (Ed.), Cognitive Theories in Social Psychology. New York : Academic Press.

ROSS, L., AMABILE, T.M. & STEINMETZ, J.L. (1977). Social roles, social control and biases in social perception processes. Journal of Personality and Social Psychology, 35, 385-394.

ROSS, L. & ANDERSON, C. (1980). Shortcomings in the attribution process : On the origins and maintainance of erroneous social judgement. In A.Tversky and D. Kahneman & P. Slovic (Eds.). Judgement Under Uncertainty : Heuristics and Biases. New York : Cambridge University Press.

ROSS, L., BIERBRAUER, G. & POLLY, S. (1974). Attribution of educational outcomes by professorial and non-professorial instructors. Journal of Experimental Social Psychology, 29, 609-618.

ROSS, L., GREENE, D. & HOUSE, P. (1977). The "false consensus effect" : An egocentric bias in social perception and attribution processes. Journal of Experimental and Social Psychology, 13, 279-301.

ROSS, L., LEPPER, M.R. & HUBBARD, M. (1979). Perseverence in self-perception : Biased attributional processes in the debriefing paradigm. Journal of Personality and Social Psychology, 32, 880-892.

ROSS, L., LEPPER, M.R., STRACK, F. & STEINMETZ, J.L. (1971). Social explanation and social expectation : The effects of real and hypothetical explanations on subjective likelihood. Journal of Personality and Social Psychology, 35, 817-829.

ROTHBART, M., FULERO, S., JENSEN, C., HOWARD, J. & BIRRELL, P. (1978). From individual to group impressions : Availability heuristics in stereotype formation. Journal of Experimental Social Psychology, 14, 237-255.

SCHANK, R. & ABELSON, R.P. (1977). Scripts, Plans, Goals and Understanding : An inquiry into Human Knowledge Structures. Hillsdale, New Jersey : Lawrence Erlbaum.

SCHEFF, T.J. (1974). The labelling theory of mental illness. American Sociological Review, 39, 444-457.

SCHNEIDER, D.J., HASTORF, A.H. & ELLSWORTH, P.C. (1979). Person Perception. London : Addison-Wesley Publishing Company.

SCHULTZ, W.R. (1967). The Phenomonology of the Social World. Evanston, Illinois : Northwestern University Press.

SCHUM, D.A. (1977). Contract effects in inference : On the conditioning of current evidence by prior evidence. Organisational Behaviour and Human Performance, 18, 217-253.

SEALY, A.P. (1975). The jury : Decision making in a small group. In H. Brown & R. Stevens (Eds.). Social Behaviour and Experience. Hodder and Stoughton Ltd.

SEALY, A.P. & CORNISH, W.R. (1973). Jurors and their verdicts. Criminal Law Review, 10, 208-223.

SEVON, G. (1975). Probability estimates of social events as related to values. Scandinavian Journal of Psychology, 16, 229-232.

SHAKLEE, H. & FISCHHOFF, B. (1979). Strategies of information search in causal analysis. Decision Research Report.

SHANTEAU, J.C. (1970). An additive decision model for sequential estimation and inference judgements. Journal of Experimental Psychology, 85, 181-191.

SHERIF, M. & HOVLAND, C.I. (1961). Social Judgement. New Haven : Yale University Press.

SIEGEL, S. (1956). Non-parametric statistics for the Behavioural Sciences. Tokyo : McGraw-Hill, Kogaburka Ltd.

SIMON, H.A. (1957a). A behavioural model of rational choice. Quarterly Journal of Economics, 69. 99-118

SIMON, H.A. (1957b). Models of Man : Social and National. New York : Wiley.

SIMMONS, R.G., KLEIN, S.D. & THORNTON, K. (19). The family member's decision to be a kidney transplant donor. Journal of Comparative Family Studies, 4, 88-115.

SLOVIC, P. (1972). From Shakespeare to Simon : Speculations - and some evidence - about man's ability to process information. Oregan Research Institute, Research Bulletin, Vol. 12, No. 2 April.

SLOVIC, P. (1963). Value as a determiner of subjective probability. I.E.E.E. Transactions on Human Factors in Electronics, Vol. HFE-7 22-28.

SLOVIC. P. & FISCHHOFF, B. (1979). How safe is safe enough ? : Determinants of perceived and acceptable risk. In L. Gould & C.A. Walker, (Eds.). Too Hot to Handle : Social and Policy issues in the Management of Radioactive Waste. Yale University Press.

SLOVIC, P. & FISCHHOFF, B. (1977). On the psychology of experimental surprises. Journal of Experimental Psychology : Human Perception and Performance, 3, 544-551.

SLOVIC, P., FISCHHOFF, B. & LICHTENSTEIN, S. (1977). Behavioural decision theory. Annual Review of Psychology, 28, 1-39.

SLOVIC, P., KUNREUTHER, H. & WHITE, G.F. (1974). Decision processes, rationality and adjustment to natural hazards. In G.F. White (Ed.), National Hazards : Local, National and Global. New York : Oxford University Press.

SLOVIC, P. & LICHTENSTEIN, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgement. Organisational Behaviour and Human Performance, 6, 649-744.

SMART, R. (1966). Subject selection bias in psychological research. Canadian Psychologist, 7, 115-121.

SMEDSLUND, J. (1963). The concept of correlation in adults. Scandinavian Journal of Psychology, 4, 165-173.

SNYDER, M. (in Press). Seek and ye shall find : Testing hypotheses about other people. In E.T. Higgins, C.P. Herman and M.P. Zanna, (Eds.) Social Cognition : The Ontario Symposium on Personality and Social Psychology. Hillsdale, New Jersey : Lawrence Erlbaum Associates.

SNYDER, M. & CANTOR, N. (1979). Testing hypotheses about other people : The use of historical knowledge. Journal of Experimental Social Psychology. 15, 330-342.

SNYDER, M. & SWANN, W.B. (1978a). Behavioural conformation in social interaction : From social perception to social reality. Journal of Experimental Social Psychology. 14, 148-162.

SNYDER, M. & SWANN, W.B. (1978b). Hypothesis-testing processes in social interaction. Journal of Personality and Social Psychology. 36, 1202-1212.

SNYDER, M. & URANOWITZ, S.W. (1978). Reconstructing the past : Some cognitive consequences of person perception. Journal of Personality. 44, 501-517.

SPONBERG, H. (1946). A study of the relative effectiveness of climax and anticlimax order in argumentative speech. Speech Monographs. 13, 35-44.

STEINBRUNER, J.D. (1974). The Cybernetic Theory of Decision : New Dimensions of Political Analysis. Princetown, New Jersey : Princetown University Press.

STONE, V.A. (1969). A primacy effect in decision making by jurors. The Journal of Communications. 19, 239-247.

STREUFERT, S. & STREUFERT, S.C. (1969). Effects of conceptual structure, failure and success on attribution of causality and interpersonal attitudes. Journal of Personality and Social Psychology. 11, 138-147.

SUE, S., SMITH, R.E. & CALDWELL, C. (1973). Effects of inadmissable evidence on the decisions of simulated jurors : A moral dilemma. Journal of Applied Social Psychology. 3, 345-353.

TAPP, J.L. (1977). Psychology and the law : A look at the interface. In B.D. Sales, (Ed.), Psychology and the legal process. Spectrum, New York : Halstead Press.

TAPP, J.L. (1970). Psychology and the law : An overture. Annual Review of Psychology. 359-404.

TAYLOR, S.E. & FISKE, S.T. (1978). Salience, attention and attribution : Top of the head phenomena. In L. Berkowitz, (Ed.), Advances in Experimental Social Psychology (Vol. II). New York : Academic Press.

TAYLOR, S.E. & FISKE, S.T. (1975). Point of view and perception of causality. Journal of Personality and Social Psychology. 32, 439-445.

THIBAUT, J. & WALKER, L. (1975). Procedural Justice : A Psychological Analysis. Hillsdale, New Jersey : Lawrence Erlbaum Associates.

THURSTONE, L.L. (1959). The Measurement of Values. Chicago : University of Chicago Press.

THURSTONE, L.L. & CHAVE, E.J. (1929). The Measurement of Attitudes. London : The University of Chicago Press.

TIETS, N.W. (1965). Comparative study of immunologic and biologic pregnancy tests in early pregnancy. Obstetrics and Gynaecology. 46, 488-499.

TUCKER, D.H. & ROW, P.M. (1979). Relationship between expectancy, causal attributions and final hiring decisions in the employment interview. Journal of Applied Psychology. 64, 27-34.

TULVING, E. & THOMSON, D.H. (1973). Encoding specificity and retrieval processes in episodic memory. Psychological Review. 80, 352-373.

TULVING, E. & WATKINS, M.J. (1975). Structure of memory traces. Psychological Review. 82, 261-275.

TVERSKY, A. (1972). Elimination by aspects : A theory of choice. Psychological Review. 79, 281-299.

TVERSKY, A. (1977). Features of similarity. Psychological Review. 84, 327-352.

TVERSKY, A. & KAHNEMAN, D. (1971). Belief in the law of small numbers. Psychological Bulletin. 76, 105-110.

TVERSKY, A. & KAHNEMAN, D. (1973). Availability : A heuristic for judging frequency and probability. Cognitive Psychology. 5, 207-232.

TVERSKY, A. & KAHNEMAN, D. (1974). Judgement under uncertainty : Heuristics and biases. Science. 185, 1124-1131.

TVERSKY, A. & KAHNEMAN, D. (1977). Causal Schemata in judgement under uncertainty. In M. Fishbein, (Ed.). Progress in Social Psychology. Hillsdale, New Jersey : Lawrence Erlbaum Ass.

TVERSKY, A. KAHNEMAN, D. & SLOVIC, P. (1980). Judgement Under Uncertainty : Heuristics and Biases. New York : Cambridge University Press.

WALKER, L., THIBAUT, J. & ANDREOLI, V. (1972). Order of presentation at trial. Yale Law Journal. 82, 216-226.

WALLACE, W. & WILSON, W. (1969). Reliable recency effects. Psychological Reports. 25, 311-317.

WALSTER, E. (1966). The assignment of responsibility for an accident. Journal of Personality and Social Psychology. 5, 508-516.

WALSTER, E. (1967). Second guessing important events. Human Relations. 20, 239-250.

WASON, P.C. & JOHNSON-LAIRD, P.N. (1965). Psychology of Reasoning : Structure and Content. London : Batsford.

WEARY, G. (1979). Self-serving attributional biases : Perceptual or response distortions. Journal of Personality and Social Psychology, 37, 1418-1420.

WELLS, G.L. & HARVEY, J.H. (1977). Do people use consensus information in making causal attributions. Journal of personality and Social Psychology, 35, 279-293.

WILSON, W. (1971). Source credibility and order effects. Psychological Reports, 29, 1303-1312.

WOHLSTETTER, R. (1965). Cuba and Pearl Harbour : Hindsight and Foresight. Foreign Affairs, 691-707.

WOHLSTETTER, R. (1962). Pearl Harbour : Warning and Decision. Stanford, California : Stanford University Press.

WOLF, S. & MONTGOMERY, D.A. (1977). Effects of inadmissable evidence and level of judicial admonishment to disregard on the judgement of mock jurors. Journal of Applied Social Psychology, 7, 205-219.

WOOD, G. (1978). The knew-it-all-along effect. Journal of Experimental Psychology : Human Perception and Performance, 4, 345-353.

WYER, R.S. (1977). The role of logical and non-logical factors in making inferences about category membership. Journal of Experimental Social Psychology, 13, 577-595.

YOUNG, F. & CLIFF, N. (1971). Interactive scaling with individual subjects. The L.L. Thurstone Psychometric Laboratory Report, No. 94, Chapel Hill, N.C.

ZADNY, J. & GERARD, H.B. (1974). Attributed intentions and information selectivity. Journal of Experimental Social Psychology, 10, 34-52.

ZAJONC, R.B. (1960). The concepts of balance, congruity and dissonance. The Public Opinion Quarterly, 24, 280-296.

ZUCKERMAN, M. (1978). The use of consencus information in prediction of behaviour. Journal of Experimental Psychology, 163-171.

A P P E N D I C E S

APPENDIX 5A - STATEMENTS FROM THIBAUT & WALKER (1975) USED IN JURY CASE (A) AND (B) IN EXPERIMENT 1

FACTS	SCALE VALUE
Adams (the defendant) and Zemp had been close friends for years. Recently, they began to gamble heavily, and, as matters became involved, met at a tavern to discuss their relationship.	NA[1]
After a period of conversation Zemp knocked Adams to the floor and threw an object in his direction.	NA
Adams responded by stabbing Zemp in the stomach with a piece of glass.	NA
The law provides that it is unlawful to use more force in repelling an attack than a person believes necessary.	NA
At the start of the meeting Adams told Zemp they should end their relationship before serious trouble developed between them.	9.00[2]
A waiter heard the defendant fall to the floor and scream "I ought to kill you for that".	3.75
The defendant ran out of the back door after the stabbing.	1.75
The defendant (Adams) is 6 foot tall and weighs 15 stones; Zemp is 5 foot 8 inches tall and weighs 12 stones.	4.50
Adams was not injured when he was knocked to the floor.	3.50
Zemp liked to drink heavily and often became violent when drunk.	10.50
Adams drank only one beer on the night of the stabbing.	L4.52[3]
Before he met Zemp at the tavern the defendant told a friend he hoped that he could settle his dispute with Zemp amiably.	11.50

[1] Statements with NA as scale value indicate those that formed part of the "test case summary", for which Thibaut & Walker did not determine scale values.

[2] Larger values indicate greater degrees of lawfulness, lower values greater degree unlawful.

[3] Value obtained from different scaling procedure by Thibaut & Walker. Letter "L" indicates Lawful classification. Larger number greater degree of lawfulness.

APPENDIX 5A (cont.)

FACTS	SCALE VALUE
One of the waiters told the police he wouldn't mind if Adams never came to the tavern again because he sometimes bullied other customers.	4.50
Zemp told one of the waiters at the tavern he planned to settle his differences with the defendant once and for all.	11.00

APPENDIX 5B - SUMMARY OF MEDIAN PROBABILITY ESTIMATES GIVEN AFTER EACH ITEM OF INFORMATION FOR EACH OUTCOME AND FOR EACH OF THE FOUR SCENARIOS

1. Mrs. Dewar in Therapy

OUTCOME	CONDITION	STATEMENTS												H1
		1	2	3	4	5	6	7	8	9	10	11	12	
1.TERMINATED THERAPY - NO IMPROVEMENT	F	5	7.5	4	1.5	2.5	5	7.5	2.5	3.5	5	2.5	2.5	2.5
	H2	2.5	5	5	5	0	5	2.5	0	0	0	0	0	
2.TERMINATED THERAPY - IMPROVEMENT	F	2.5	2.5	2.5	2.5	5	2.5	2.5	2.5	2.5	2.5	2.5	5	0
	H2	2.5	0	2.5	7.5	0	2.5	0	0	0	0	2.5	1	
3.CONTINUED THERAPY - NO IMPROVEMENT	F	20	10	10	10	7.5	10	12.5	10	12.5	15	17.5	12.5	7.5
	H2	50	35	37.5	37.5	17.5	35	25	27.5	45	60	47.5	30	
4.CONTINUED THERAPY - IMPROVEMENT	F	60	62.5	62.5	72.5	72.5	70	75	77.5	77.5	77.5	72.5	70	82.5
	H2	22.5	40	50	50	57.5	45	57.5	60	52.5	40	50	55	

APPENDIX 5B - SUMMARY OF MEDIAN PROBABILITY ESTIMATES GIVEN AFTER EACH ITEM OF INFORMATION FOR EACH OUTCOME AND FOR EACH OF THE FOUR SCENARIOS

2. The British-Gurkha Struggle

OUTCOME	CONDITION	STATEMENTS												H1
		1	2	3	4	5	6	7	8	9	10	11	12	
1.BRITISH VICTORY	F	45	35	30	47.5	47.5	50	37.5	50	37.5	37.5	40	17.5	57.5
	H2	30	17.5	7.5	70	12.5	52.5	37.5	50	25	25	25	55	
2.GURKHA VICTORY	F	17.5	20	20	22.5	25	17.5	22.5	22.5	20	25	22.5	20	5
	H2	5	17.5	20	10	2.5	10	30	20	22.5	37.5	22.5	2.5	
3.STALEMATE - NO PEACE SETTLEMENT	F	15	20	27.5	22.5	27.5	25	25	17.5	20	20	27.5	25	0
	H2	17.5	25	37.5	0	40	12.5	37.5	2.5	42.5	25	8.5	0	
4.STALEMATE - PEACE SETTLEMENT	F	25	20	17.5	12.5	10	15	11	10	10	12.5	12.5	12.5	7.5
	H2	20	25	20	0	0	0	0	2.5	0	0	10	0	

APPENDIX 5B - SUMMARY OF MEDIAN PROBABILITY ESTIMATES GIVEN AFTER EACH ITEM OF INFORMATION FOR EACH OUTCOME AND FOR EACH OF THE FOUR SCENARIOS

3. **Jury Case A**

OUTCOME	CONDITION	STATEMENTS														H1
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	
1.NOT GUILTY	F	25	35	32.5	25	0	0	0	0	0	0	0	0	0	0	0
	H2	25	27.5	15	10	0	0	0	0	0	0	0	0	0	0	
2.GUILTY - GRIEVOUS BODILY HARM	F	25	22.5	10	27.5	75	75	55	67.5	75	65	60	50	60	37.5	25
	H2	17.5	7.5	5	22.5	25	45	29	35	25	25	25	25	15	17.5	
3.GUILTY - MORE FORCE THAN NECESSARY	F	15.5	22	28	25.5	23.5	19.5	36.5	38.0	35.5	39.5	36.5	40	42	43.5	65.5
	H2	25	40	50	35	45	45	50	47.5	50	50	50	51.5	59.5	52	
4.HUNG JURY	F	25	25	7.5	5	0	0	0	0	0	0	0	2.5	0	0	0
	H2	22.5	18.5	4	4	7.5	2.5	2.5	0	0	0	3.5	3.0	7.5	7.5	

APPENDIX 5B - SUMMARY OF MEDIAN PROBABILITY ESTIMATES GIVEN AFTER EACH ITEM OF INFORMATION FOR EACH OUTCOME AND FOR EACH OF THE FOUR SCENARIOS

4. Jury Case B

OUTCOME	CONDITION	STATEMENTS														H1
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	
1.NOT GUILTY	F	23	35	41.5	50	47.5	45	45	3	1.5	1	0.5	0.5	0	0	0
	H2	37.5	50	37.5	27.5	55	37.5	15	0	0	0	0	0	0	0	
2.GUILTY - GRIEVOUS BODILY HARM	F	5.5	5	7.5	2.5	2.5	1.25	6.5	45	55	60	30	40	65	87.5	55
	H2	25	2.5	17.5	25	10	10	17.5	50	40	40	45	35	37.5	30	
3.GUILTY - MORE FORCE THAN NECESSARY	F	13	15	7.5	25	7.5	25	34	25	21.5	29	25	17.5	15	3.5	30
	H2	25	22.5	25	25	17.5	17.5	30	37.5	50	55	45	47.5	47.5	55	
4.HUNG JURY	F	13	16	11.5	0	0	0	0	0	0	0	0	0	0	0	0
	H2	12.5	10	10	25	25	0	0	0	0	0	0	0	0	0	0

APPENDIX 6A - FISCHHOFF'S (1975) INSTRUCTIONS TO SUBJECTS IN EXPERIMENT 1

(which investigates creeping determinism)

The instructions were as follows:

> "In this questionnaire we are interested in knowing how people judge the likelihood of possible outcomes of social events. A passage describing an unfamiliar historical event appears below. We will ask you to evaluate the probability of occurrence of each of the four possible outcomes of the event in the light of the information appearing in the passage." (p.289)

Following the passage subjects were asked:

> "In the light of the information appearing in the passage, what was the probability of occurence of each of the four possible outcomes listed below? (The probabilities should sum to 100%)" (p.289)

APPENDIX 6B - SUMMARY OF MANN WHITNEY U VALUES AND SIGNIFICANCE LEVELS FOR FORESIGHT/HINDSIGHT COMPARISONS

SCENARIO AND OUTCOME NUMBER		COMPARISON	U	SIGNIFICANCE
MRS.DEWAR IN THERAPY	1	F - H(B)	31.5	NS
	1	F - H(A)	44.5	NS
	2	F - H(B)	47	NS
	2	F - H(A)	42.5	NS
	3	F - H(B)	30.5	NS
	3	F - H(A)	32.5	NS
	4	F - H(B)	50	NS
	4	F - H(A)	39	NS
THE BRITISH - GURKHA STRUGGLE	1	F - H(B)	47	NS
	1	F - H(A)	19.5	PLO.025
	2	F - H(B)	36.5	NS
	2	F - H(A)	28.5	PLO.1
	3	F - H(B)	38	NS
	3	F - H(A)	30.5	NS
	4	F - H(B)	49.5	NS
	4	F - H(A)	40	NS
JURY CASE A	1	F - H(B)	48	NS
	1	F - H(A)	40.5	NS
	2	F - H(B)	47.5	NS
	2	F - H(A)	32.5	NS
	3	F - H(B)	36.5	NS
	3	F - H(A)	32.5	NS
	4	F - H(B)	27	PLO.05
	4	F - H(A)	43.5	NS
JURY CASE B	1	F - H(B)	48.5	NS
	1	F - H(A)	39	NS
	2	F - H(B)	27	PLO.05
	2	F - H(A)	37.5	NS
	3	F - H(B)	29	PLO.1
	3	F - H(A)	40.5	NS
	4	F - H(B)	40.5	NS
	4	F - H(A)	44.5	NS

F = foresight; H(A) = hindsight after; H(B) = hindsight before

APPENDIX 7A - QUESTIONNAIRE FOR "MRS.S" IN FORESIGHT ORDER 1

QUESTIONNAIRE

Sex .....

(1) In the light of what you have just read would you indicate how probable you think each of the four following outcomes are within the 5 year period following Mrs. S.'s third suicide attempt. Please give your probability estimates in percentages and ensure that the four scores add up to 100%.

Outcome	Probability of Occurrence
1. Successful suicide attempt.	
2. Further suicide attempt regarded as being highly dangerous to life and of serious intent on her part.	
3. Further suicide attempt regarded as not being dangerous to life or of serious intent on her part.	
4. No further suicide attempt.	

(2) Again in the light of what you have read would you please indicate for each outcome how surprised you would be if it occurred. Would you do this on the 7-point scale provided.

1. Successful suicide attempt.

7 6 5 4 3 2 1

Surprised ... Not Surprised

2. Further suicide attempt regarded as being highly dangerous to life and of serious intent on her part.

7 6 5 4 3 2 1

Surprised ... Not Surprised

3. Further suicide attempt regarded as not being dangerous to life or of serious intent on her part.

7 6 5 4 3 2 1

Surprised ... Not Surprised

4. No further suicide attempt.

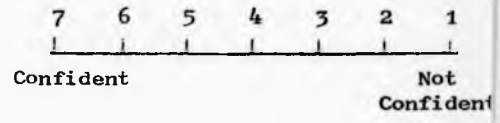
7 6 5 4 3 2 1

Surprised ... Not Surprised

APPENDIX 7A (cont.)

Questionnaire continued

(3) Would you please indicate how confident you feel about the judgements you have just made concerning Mrs.S.



APPENDIX 7B - QUESTIONNAIRE FOR "MRS.S" IN HINDSIGHT ORDER 1

QUESTIONNAIRE

Sex .....

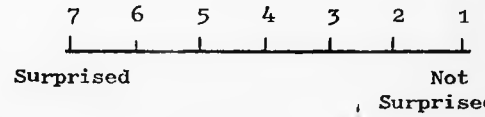
(1) In the light of what you have just read would you indicate how likely you think each of the four outcomes listed below were in the 5 year period following Mrs.S.'s third suicide attempt. Please give your probability estimates in terms of percentages and ensure that the four scores add up to 100%.

Outcome	Probability of Occurrence
1. Successful suicide attempt*	
2. Further suicide attempt which was regarded as being highly dangerous to life and of serious intent on her part.	
3. Further suicide attempt which was regarded as not being dangerous to life or of serious intent on her part.	
4. No further suicide attempt.	

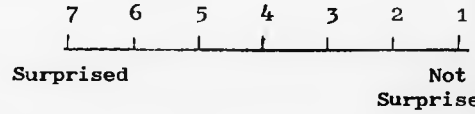
*This is the outcome which actually happened.

(2) Again in the light of what you have just read would you please indicate for each outcome how surprised you would have been had it occurred. Would you do this on the 7-point scale provided.

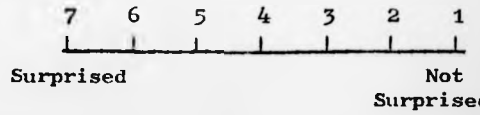
1. Successful suicide attempt.



2. Further suicide attempt which was regarded as being highly dangerous to life and of serious intent on her part.



3. Further suicide attempt which was regarded as not being dangerous to life or of serious intent on her part.



APPENDIX 7B (cont.)

Questionnaire continued

4. No further suicide attempt.

7 6 5 4 3 2 1

Surprised — Not Surprised

(3) Would you please indicate how confident you feel about the judgements you have made concerning Mrs.S.

7 6 5 4 3 2 1

Confident — Not Confident

APPENDIX 7C - FORESIGHT ORDER 1 FOR MRS.S SCENARIO

MRS. S.

EVENT DESCRIPTION

Mrs. S. loved her three children and appeared to get a sense of fulfillment from them. A short while after her husband's death Mrs. S. returned to work as a secretary and seemed to enjoy going back to work. Mrs. S. was a likeable person who always appeared to be cheerful and had a free and easy attitude to life. She had left home at the age of 21 and took a job as a secretary at which she was quite competent and which she seemed to find enjoyable. At 29 Mrs. S. was married, she gave up her secretarial work to have a family. In a period of 7 years, from 30 to 37, Mrs. S. had three children - two girls and a boy. Mrs. S. had had a fairly happy childhood and did not want for any material things. Insurances had left her well provided for materially and financially after her husband's death. Mrs. S. had no previous psychiatric history of mental illness apart from that associated with her second suicide attempt. Mrs. S. was not altogether happy with her life as a housewife, it often made her tense and depressed. At the age of 20 she had made her first suicide attempt by trying to poison herself with her mother's sedatives - this was classified as a dangerous threat to life but not of serious intent on her part. Mrs. S. was born in 1930 and came from a family in which her father had committed suicide and her mother often threatened it. After her first suicide attempt Mrs. S. was restless and discontented perhaps because she did not have a steady boyfriend. When she was 40 her husband died of a heart attack which made her depressed and discontented. A year after her husband's death she made a third suicide attempt by taking a small overdose of aspirins - this was seen as relatively harmless to life and not of serious intent on her part. At 27 Mrs. S. had made a second suicide attempt by poisoning - this was seen as dangerous to life and of serious intent on her part - she was admitted to a mental hospital for two months.

APPENDIX 7D - JUDGES' SCALES FOR MRS.S. AND ADAMS AND ZEMP SCENARIOS

SCALES

(i) MRS.S.

1	2	3	4	5	6	7	8	9
weakly indicate no further suicide attempt								strongly indicate no further suicide attempt

1	2	3	4	5	6	7	8	9
weakly indicate a further suicide attempt								strongly indicate a further suicide attempt

(ii) ADAMS AND ZEMP

1	2	3	4	5	6	7	8	9
lowest degree of guilt								highest degree of guilt

1	2	3	4	5	6	7	8	9
lowest degree of innocence								highest degree of innocence

APPENDIX 7E - CUMULATIVE FREQUENCIES AND SCALE VALUES FOR EACH SENTENCE IN THE "MRS.S" SCENARIO

Mrs. S. - Summary of scaling of statements by 20 persons

Statement Number	Scale (+ or -)	Scale Value	CUMULATIVE FREQUENCY																
			-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9
1	+	5.0	100	100	100	100	100	100	100	100	100	100	85	65	65	60	50	50	45	35
2	-	4.5	0	10	20	35	45	55	65	75	90	90	95	95	95	100	100	100	100	100
3	+	3.3	100	100	100	100	100	100	100	100	100	100	75	55	35	30	20	5	5	5
4	-	4.2	5	5	20	30	35	55	60	80	85	90	90	95	95	100	100	100	100	100
5	-	5.0	5	10	20	40	50	60	60	75	95	95	95	100	100	100	100	100	100	100
6	+	5.0	100	100	100	100	100	100	100	100	100	100	90	75	65	50	40	25	15	5
7	-	7.7	30	45	60	75	80	90	90	90	100	100	100	100	100	100	100	100	100	100
8	+	4.7	100	100	100	100	100	100	100	100	100	95	80	70	60	45	35	15	10	10
9	+	4.3	100	100	100	100	100	100	100	100	95	95	75	65	55	30	20	15	15	10
10	-	3.2	0	10	15	15	30	35	55	75	60	100	100	100	100	100	100	100	100	100
11	-	5.7	5	35	40	45	60	70	75	75	95	95	100	100	100	100	100	100	100	100
12	+	2.5	100	100	100	100	100	100	100	100	100	95	60	40	20	10	10	5	5	5
13	+	5.0	100	100	100	100	100	100	100	100	100	100	95	80	65	50	40	20	5	5
14	-	6.0	5	5	30	50	55	65	70	75	100	100	100	100	100	100	100	100	100	100
15	+	7.4	100	100	100	100	100	100	100	100	100	100	100	90	80	75	60	55	35	20
16	+	1.3	100	100	100	100	100	90	85	80	75	60	25	10	5	5	5	0	0	0

(-) - a further suicide attempt

(+) - no further suicide attempt

APPENDIX 7F - CUMULATIVE FREQUENCIES AND SCALE VALUES FOR EACH SENTENCE IN THE "ADAMS & ZEMP" SCENARIO

Adams & Zemp - Summary of scaling of statements by 20 persons

Statement Number	Scale (+ or -)	Scale Value	CUMULATIVE FREQUENCY																	
			-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9
1	+	6.0	100	95	95	95	95	95	95	90	85	80	65	65	60	60	50	40	25	20
2	+	2.0	100	100	100	100	100	100	95	90	80	60	50	45	40	40	15	15	10	10
3	+	4.3	100	100	100	100	95	95	85	80	75	70	65	60	55	40	35	30	25	10
4	+	5.7	100	100	100	100	100	100	100	95	95	70	70	60	60	60	45	30	20	20
5	-	8.0	40	50	55	75	85	90	95	95	100	100	100	100	100	100	100	100	100	100
6	-	6.4	15	30	40	55	65	65	75	75	85	90	95	100	100	100	100	100	100	100
7	-	5.4	10	25	35	45	55	65	70	75	85	90	95	95	95	100	100	100	100	100
8	-	5.3	5	25	35	35	55	65	70	70	75	85	90	90	90	95	95	95	95	100
9	+	5.0	100	100	100	95	95	90	90	90	85	75	70	60	55	50	20	15	5	5
10	-	5.4	10	25	35	40	60	60	75	80	85	90	95	100	100	100	100	100	100	100
11	+	2.0	100	100	95	95	95	90	85	70	70	60	50	45	40	35	30	20	10	5
12	+	7.4	100	100	100	100	100	100	100	100	100	90	90	85	75	75	55	55	35	10
13	-	4.4	0	5	15	35	40	55	60	65	70	75	75	80	85	90	90	95	95	100
14	+	4.5	100	100	100	95	90	90	90	90	90	90	90	85	60	40	35	30	15	10

(-) - Guilty

(+) - Not Guilty

APPENDIX 8A - QUESTIONNAIRE GIVEN TO "JURORS"

QUESTIONNAIRE

Now that you have read the case, I would like you to reach a verdict on each of the four charges. Would you please fill in the following questionnaire indicating your verdicts and the degree of confidence you feel in each one.

(1) Would you find Harrison guilty or not guilty on the charge of rape? Please indicate by putting either G (guilty) or NG (not guilty) in the box below.

Harrison on the charge of rape ☐

Having done this would you please indicate how confident you are about your verdict. Please indicate on the following seven point scale :-

Not confident at all — Very confident

0 1 2 3 4 5 6 7

(2) Would you find Harrison guilty or not guilty on the charge of attempted rape? Please indicate by putting either G (guilty) or NG (not guilty) in the box below.

Harrison on the charge of attempted rape ☐

Having done this would you please indicate how confident you are about your verdict. Please indicate on the following seven point scale :-

Not confident at all — Very confident

0 1 2 3 4 5 6 7

(3) Would you find Bryce guilty or not guilty on the charge of rape? Please indicate by putting either G (guilty) or NG (not guilty) in the box below :-

Bryce on the charge of rape ☐

Having done this would you please indicate how confident you are about your verdict. Please indicate on the following seven point scale :-

Not confident at all — Very confident

0 1 2 3 4 5 6 7

APPENDIX 8A (cont.)

Questionnaire continued

(4) Would you find Bryce guilty or not guilty on the charge of attempted rape? Would you please indicate by putting either G (guilty) or NG (not guilty) in the box below :-

Bryce on the charge of attempted rape [ ]

Having done this would you please indicate how confident you are about your verdict. Please indicate on the following seven point scale :

Not confident at all — Very confident

0 1 2 3 4 5 6 7

APPENDIX 8B - TRIAL CASE AS READ BY SCALING JUDGES

(1)

RAPE TRIAL

PROSECUTION'S CASE

James William Harrison and John Michael Bryce are charged with the rape of Mary Watkins on November 5th at Boothville, Liverpool. Each pleads not guilty.

Prosecution says that Harrison and Bryce are charged with raping 16 year old Mary Watkins of Boothville. He tells them that she and her friend Joyce Blunden attended the bonfire on allotments near Mary Watkins' home and after talking to a group of friends they set off for home. Harrison approached them as they were talking down Vale Street, he frightened them and they ran back to their friends. Fifteen minutes later they set off again and this time they were chased by Harrison through inlets between the houses where Mary Watkins was grabbed by Harrison and taken into an entry behind the houses. It is the Crown's case that Harrison held Mary Watkins against her will and forced her to have intercourse against her will, and that Bryce came up and also forced her to have intercourse with him. You will hear (a) what the girl went through (b) Mr. George Smith, a householder, who overheard what was taking place, that they would not let Mary Watkins go when she wanted to (c) Mrs. Godfrey, who met Mary Watkins on her way home afterwards (d) Joyce Blunden who was with her at the bonfire (e) Mary Watkins mother. Prosecution asserts that all this evidence suggests that Mary Watkins did not consent to these men interfering with her sexually and that she was thoroughly upset, flushed and crying. After hearing the evidence you will come to hold the view that both defendents did (or at least attempted to) have sexual intercourse without her consent.

MARY WATKINS' EVIDENCE

Prosecutor obtains statements of the events from Mary Watkins to the effect that on Tuesday November 5th at about 10 PM she and Joyce Blunden were talking with some boys at the Vale Street bonfires when Harrison and Bryce passed them. Mary Watkins states eventually she and Joyce Blunden left the bonfire, intending to go home, and that, when they had walked a

APPENDIX 8B (cont.)

little way down Vale St., before reaching Victoria Rd., Harrison ran towards them and they ran back to the boys, but Harrison had grabbed and held Mary Watkins for a while. After 10 minutes she and Joyce Blunden set off for home again, in the same direction. Mary Watkins says that when Harrison chased them they ran down a passage between houses and that half way down she ran into Bryce who got hold of her by the arm and that Harrison then got hold of her left wrist. She next states that Harrison dragged her over waste ground into an entry where he pushed her up against a wall, undid the zip of her slacks, pulled them down, pulled down her knickers and put his hand round her waist. She says that she was not willing for Harrison to do this and that she tried to stop him but that he hit her, banging her head against the wall and punching her on the arm. She states that he tried to have intercourse with her and that he succeeded. She said "when he dragged me into the entry, he pulled down my pants, brought out his person, forced my legs open, he put his person in between my legs and I managed to break away from him."

Further questioning elicits statements from Mary Watkins to the effect that Bryce twice came into the entry to ask Harrison if he had finished. The first time Harrison is alleged to have said no, and tried again to have sexual intercourse, the second time Bryce returned and this time Harrison said he had finished.

Mary Watkins states that the two whispered, leaving her still unclothed and Bryce then thrust her against the wall and brought out his person, moving it around and asking her to touch it, which she refused. Next she alleges that Bryce tried to put one finger then two inside her, causing her to scream and that he threatened to do what his mate had to her if she screamed. He then put his person in between her legs which hurt her. She says that Bryce said "It will hurt" and that he kept on telling her to press and then pulled his person out of her. The prosecution establishes that Mary Watkins is saying that Bryce actually got inside her. Mary Watkins then states that Bryce said "you can go now, but don't tell anyone or I'll get you again". She then pulled up her clothing and then ran out into Victoria Rd.

APPENDIX 8B (cont.)

The prosecutor then requires Mary Watkins to state that being in the entry with the two men and what went on was without her consent.

On further questioning Mary Watkins states that she met Mrs. Godfrey in Vale St. and that she was crying, her nose was bleeding, her wrist bruised and her arm red. She says that Mrs. Godfrey took her to Mrs. Blunden's house. Joyce Blunden then asked Mary Watkins what had happened to her and was told that she had been stripped by the two lads. In reply to Joyce Blunden's question "did they do anything to you" Mary Watkins said "Yes".

1. Defence establishes that Harrison was kissing Mary Watkins before the alleged intercourse attempt.

2. Mary Watkins tells the defence lawyer that she knew the meaning of intercourse before November 5th.

3. On further cross questioning Mary Watkins states that Harrison was kissing her for at least three minutes, and that this occurred before Harrison unfastened her trousers.

4. Defence establishes that at the magistrates court Mary Watkins had said that her trousers had fallen down and that now she asserts that they were pulled down. She said she had got mixed up earlier and had meant to say that they were pulled down.

5. She also failed to tell the magistrates court that Harrison had been kissing her.

6. She didn't reply when asked why she had not told the court that she had been kissed by Harrison.

APPENDIX 8B (cont.)

7. Also, when giving evidence in the magistrates court she had told Joyce Blunden that two men had stripped her and that she was indicating that both men had removed her clothes. Today she said that Harrison only had stripped her. She said that she had got mixed up earlier.

8. Mary Watkins had said that Harrison had made her nose bleed, however when talking to Det.Con. Foster at the police station she had said that Bryce had done it. She was sure now that it was Harrison.

9. She admitted that she wasn't savagely attacked, but had been hit on the face and banged against the wall.

10. When giving evidence at the magistrates court she had said "when the bolt rattled Bryce was present". However, today she said that only Harrison was present when the bolt rattled.

11. Defence suggests that she went willingly into the entry with Harrison, that there was affection between them and that she let him undo her trousers. Mary Watkins denies it.

12. Defence then suggests that Bryce allowed her to go as soon as he arrived. She denies this also.

JOYCE BLUNDEN'S EVIDENCE

13. In reply to prosecution's questions Joyce Blunden states that at about 10 pm on November 5th she and Mary Watkins were talking to

APPENDIX 8B (cont.)

some boys at the corner of Victoria Rd. and Vale St. when the accused men passed by. She further states that they left the group to go home and were chased by Harrison who caught Mary Watkins and let her go. Both of them returned to the boys.

14. She then states that after a quarter of an hour they started for home again. The two men approached them and Bryce chased her, she ran home without seeing what happened to Mary Watkins.

15. The prosecutor questions Joyce Blunden about later that night and she says Mary Watkins arrived at her home with Mrs. Godfrey, that Mary Watkins was crying and her nose was bleeding.

16. In trying to elicit what was said between the two girls the prosecutor was told that Mary Watkins said she had been raped. The judge establishes the word wasn't used and that Mary Watkins had said "they stripped me".

CROSS EXAMINATION OF JOYCE BLUNDEN

17. Defence establishes that Joyce Blunden had said at the earlier hearing that she had been chased only once that night but that now she is alleging that it was twice.

GEORGE SMITH'S EVIDENCE

18. George Smith states that on the night in question while going to the outside toilet he heard a couple outside his back gate. The girl said hysterically "You've had what you wanted, can I go now?". Then he and his wife heard the girl say "I'm tired, will you let me go, I have to go to work in the morning".

APPENDIX 8B (cont.)

19. He states that his wife went indoors and that he rattled the bolt to move the couple.

20. He went in but when he came out later he heard footsteps and a man's voice saying "I've got her mate round the corner".

21. George Smith says he heard footsteps across the waste ground again, and that he went indoors again.

22. On coming out again he heard the girl say "Please let me go, I will give you anything you want". He heard the man decide to let her go and say "Don't go that way. Go that way.".

23. Then he heard a girl's footsteps going down the entry towards Lyle Rd.

CROSS EXAMINATION OF GEORGE SMITH

24. Defence establishes that both could hear the bolt rattled and that he didn't hear sounds of violence, screaming or anyone in pain.

MRS. GODFREY'S EVIDENCE

25. In reply to questioning Mrs. Godfrey states that she lives three doors from Mary Watkins' home and that on November 5th at about 11 pm, returning home from a visit to her sister's, she heard Mary Watkins call out to her in Vale St.. She states that Mary Watkins was very distressed and crying and that later at Joyce Blunden's home she saw that Mary Watkins' nose was bleeding. She then took Mary Watkins home.

APPENDIX 8B (cont.)

DOROTHY WATKINS' EVIDENCE

26. Dorothy Watkins, the mother of Mary Watkins, states that at about 11 pm on November 5th Mrs. Godfrey brought her daughter home with a red, swollen face, red upper arms, a blackened left wrist, nose bleed and crying terribly.

CROSS EXAMINATION OF DOROTHY WATKINS

27. Defence establishes that the swollen face may have been caused by crying.

28. Defence also establishes that Dorothy Watkins did not know of any boyfriends of Mary Watkins.

29. Dorothy Watkins was sure that Mary was quite innocent of what sexual relations were.

PROSECUTION READS EVIDENCE OF DR. GROSVENOR

30. Mary Watkins was examined on 6th November at 11.30 am. The doctor states that he found the condition of the hymen and vagina consistent with partial penetration by the male organ, a condition that could have been caused otherwise.

31. There was no sign of recent tearing of the orifice, but it was possible that a woman whose vagina was already in such a condition could have sexual intercourse and there be little physical evidence afterwards.

32. A vaginal swab did not reveal the presence of spermatozoa nor was any foreign hair present on the patient's body.

APPENDIX 8B (cont.)

DET. CON. FOSTER'S EVIDENCE

33. Harrison was interviewed by me in the presence of P.C. Hemingway. Initially he denied (a) that he carried darts (one of the men alleged to have raped Mary Watkins was seen to recover darts), (b) that he was connected with the affair at all (alleging he was in the Iron Lung public house). P.C. Hemingway went to investigate these claims and failed to substantiate them. On P.C. Hemingway's return Harrison produced darts from his pocket and admitted "It was me, I may as well tell you all about it".

34. After cautioning Harrison made the following statement:-

(i) "On Tuesday, I think it was after 10 o'clock, I was in Victoria Rd. with two of my mates. We were standing outside a public house. I saw two girls standing on the opposite corner of Victoria Rd. I went over to them and they were talking to some lads. I got hold of one of the girls and took her down the entry.

(ii) "She never said much so I undid her trousers, they fell down. She had a pair of knickers on and I didn't try to pull them down. I took my knob out and tried to put it between her legs but she wouldn't. I got fed up with her so I let her go, I put my knob back in my trousers and walked out of the entry.

(iii) "I don't know if anybody else went into the entry after me."

35. Bryce was seen by Det. Con. Foster on 11th November and invited to make a statement about the alleged rape. After cautioning he made the following statement:-

(i) "On bonfire night I was with a group of lads outside the public house on the corner of Victoria Rd. and Vale St. I saw Harrison go with a girl into the hollow in Victoria Rd."

APPENDIX 8B (cont.)

DET. CON. FOSTER'S EVIDENCE

33. Harrison was interviewed by me in the presence of P.C. Hemingway. Initially he denied (a) that he carried darts (one of the men alleged to have raped Mary Watkins was seen to recover darts), (b) that he was connected with the affair at all (alleging he was in the Iron Lung public house). P.C. Hemingway went to investigate these claims and failed to substantiate them. On P.C. Hemingway's return Harrison produced darts from his pocket and admitted "It was me, I may as well tell you all about it".

34. After cautioning Harrison made the following statement:-

(i) "On Tuesday, I think it was after 10 o'clock, I was in Victoria Rd. with two of my mates. We were standing outside a public house. I saw two girls standing on the opposite corner of Victoria Rd. I went over to them and they were talking to some lads. I got hold of one of the girls and took her down the entry.

(ii) "She never said much so I undid her trousers, they fell down. She had a pair of knickers on and I didn't try to pull them down. I took my knob out and tried to put it between her legs but she wouldn't. I got fed up with her so I let her go, I put my knob back in my trousers and walked out of the entry.

(iii) "I don't know if anybody else went into the entry after me."

35. Bryce was seen by Det. Con. Foster on 11th November and invited to make a statement about the alleged rape. After cautioning he made the following statement:-

(i) "On bonfire night I was with a group of lads outside the public house on the corner of Victoria Rd. and Vale St. I saw Harrison go with a girl into the hollow in Victoria Rd."

APPENDIX 8B (cont.)

DET. CON. FOSTER'S EVIDENCE

33. Harrison was interviewed by me in the presence of P.C. Hemingway. Initially he denied (a) that he carried darts (one of the men alleged to have raped Mary Watkins was seen to recover darts), (b) that he was connected with the affair at all (alleging he was in the Iron Lung public house). P.C. Hemingway went to investigate these claims and failed to substantiate them. On P.C. Hemingway's return Harrison produced darts from his pocket and admitted "It was me, I may as well tell you all about it".

34. After cautioning Harrison made the following statement:-

(i) "On Tuesday, I think it was after 10 o'clock, I was in Victoria Rd. with two of my mates. We were standing outside a public house. I saw two girls standing on the opposite corner of Victoria Rd. I went over to them and they were talking to some lads. I got hold of one of the girls and took her down the entry.

(ii) "She never said much so I undid her trousers, they fell down. She had a pair of knickers on and I didn't try to pull them down. I took my knob out and tried to put it between her legs but she wouldn't. I got fed up with her so I let her go, I put my knob back in my trousers and walked out of the entry.

(iii) "I don't know if anybody else went into the entry after me."

35. Bryce was seen by Det. Con. Foster on 11th November and invited to make a statement about the alleged rape. After cautioning he made the following statement:-

(i) "On bonfire night I was with a group of lads outside the public house on the corner of Victoria Rd. and Vale St. I saw Harrison go with a girl into the hollow in Victoria Rd."

APPENDIX 8B (cont.)

(ii) "Sometime later I went to the hollow to make water. As I got there Harrison was just leaving the girl."

(iii) "I spoke to the girl and asked her what was wrong, but she never said nothing, but appeared upset. I said something like 'you had better go home' to her and she ran down the entry. That was the last I saw of her."

CROSS EXAMINATION OF DET. CON. FOSTER

36. Defence established that Mary Watkins never accused Harrison of making her nose bleed.

37. Defence also established that Bryce had always denied having anything to do with the girl.

DEFENCE'S CASE

EVIDENCE OF HARRISON

38. In reply to defence questions Harrison states that on November 5th he met Mary Watkins when he and others were coming down the street, and that he spoke to her and there were a lot of lads on the corner. He says that he did not chase her but that there was a lot of milling about.

39. Later he says he went with her to the entry and they were kissing for three or four minutes during which time she hadn't pushed him away or said no.

40. Asked whether anyone else was there, he states that Mary Watkins had said "as long as Bryce doesn't come". He also states that he asked her where she lived and other such questions.

APPENDIX 8B (cont.)

41. Harrison states that Bryce and some others were on the bomb debris.

42. He states that he then went home and that Mary Watkins had not objected to what he did to her. He also says that he didn't see Bryce again.

CROSS EXAMINATION OF HARRISON

43. In answer to prosecution's questions Harrison states that he approached Mary Watkins (not with Bryce) when she was with a crowd of boys and put his arm on her shoulder and was speaking to her.

44. She was a stranger to him.

45. Harrison says that she soon came with him when he asked. He denies ever holding her by the wrist.

46. Prosecution reminds that in his statement he said "I got hold of one of the girls and took her down the entry."

47. He asserts that everything that took place was with her consent but that he never "got right in her" or even partly in because some lads came along.

48. Harrison asks what "penetration" means and says that when he told the judge previously "I had sexual intercourse with her" he had meant "got it between her legs". Later he says that he does not

APPENDIX 8B (cont.)

41. Harrison states that Bryce and some others were on the bomb debris.

42. He states that he then went home and that Mary Watkins had not objected to what he did to her. He also says that he didn't see Bryce again.

CROSS EXAMINATION OF HARRISON

43. In answer to prosecution's questions Harrison states that he approached Mary Watkins (not with Bryce) when she was with a crowd of boys and put his arm on her shoulder and was speaking to her.

44. She was a stranger to him.

45. Harrison says that she soon came with him when he asked. He denies ever holding her by the wrist.

46. Prosecution reminds that in his statement he said "I got hold of one of the girls and took her down the entry."

47. He asserts that everything that took place was with her consent but that he never "got right in her" or even partly in because some lads came along.

48. Harrison asks what "penetration" means and says that when he told the judge previously "I had sexual intercourse with her" he had meant "got it between her legs". Later he says that he does not

APPENDIX 8B (cont.)

think that sexual intercourse is just putting your person between a girl's legs, but that it is "you have sexual intercourse with someone, and somebody has a baby, I know that much".

49. To further prosecution questioning he says that the reason he'd said he'd had sexual intercourse with her when he'd only had it between her legs was that he thought it was the same thing.

50. Prosecution repeats his question about Mary Watkins resisting. Harrison asserts that she didn't except when the crowd came.

51. He is offered his police statement but says he can't read. He asserts that he tried on two separate occasions to get it between her legs and succeeded once.

52. Prosecution persists with questions about Mary Watkin's alleged consent and Harrison says that he asked her to open her legs. Although she said no she did in fact open them.

53. Harrison agrees that he has not been telling the truth about this. He agrees with prosecution that he tried to force her legs open and that it was not with her consent.

54. He denies that he tried to rape the girl, but gives no answer when asked what he calls what he, in fact, said he did.

EVIDENCE OF BRYCE

55. In reply to defence questions Bryce says that on the night he went across the waste ground in Victoria St. "to make water". He saw

APPENDIX 8B (cont.)

some girl there. He asked her what was up and told her she ought to go home that way so as to avoid the lads on the corner.

56. He says that she ran down the entry and that he went back on the corner with the lads. He says that he did not have sexual intercourse with her.

CROSS EXAMINATION OF BRYCE

57. In reply to prosecution questioning he says that he had seen Mary Watkins earlier in the evening but that she hadn't been running away.

58. He admits chasing and catching Joyce Blunden, saying it was part of the general "skylarking about" of bonfire night.

59. He says he let Joyce Blunden go when she said she wanted to go home and asserts he never had hold of Mary Watkins.

60. He denies saying to Harrison "have you finished yet".

61. After persistent questioning Bryce still asserts that he went into the entry only once, that he found Harrison and the girl leaving, that the girl was fully clothed but distressed, that the only thing he said to her was about which way to go out of the entry.

62. He denies successively each statement of the girl's about what she

APPENDIX 8B (cont.)

some girl there. He asked her what was up and told her she ought to go home that way so as to avoid the lads on the corner.

56. He says that she ran down the entry and that he went back on the corner with the lads. He says that he did not have sexual intercourse with her.

CROSS EXAMINATION OF BRYCE

57. In reply to prosecution questioning he says that he had seen Mary Watkins earlier in the evening but that she hadn't been running away.

58. He admits chasing and catching Joyce Blunden, saying it was part of the general "skylarking about" of bonfire night.

59. He says he let Joyce Blunden go when she said she wanted to go home and asserts he never had hold of Mary Watkins.

60. He denies saying to Harrison "have you finished yet".

61. After persistent questioning Bryce still asserts that he went into the entry only once, that he found Harrison and the girl leaving, that the girl was fully clothed but distressed, that the only thing he said to her was about which way to go out of the entry.

62. He denies successively each statement of the girl's about what she

APPENDIX 8B (cont.)

alleges he said or did to her and asserts that that part of the evidence given by the girl is a complete fabrication.

63. Bryce says at one point "No, I told her I would have nothing to do with her in that respect" but immediately asserts it is a slip of the tongue and he had meant to say "I told you, sir, I had nothing to do with her in that respect".

APPENDIX 8B (cont.)

PROSECUTION'S SUMMING UP

Prosecution asserts that both accused are guilty and that in Harrison's case there is no thought of Mary Watkins' suffering during or after the act. He points out that Harrison admitted to sexual intercourse or at least putting his penis between her legs. Prosecution says Mr. Smith's statement is evidence that two men were involved in holding a girl in the entry against her will and that Mrs. Godfrey, Joyce Blunden and Mrs. Watkins' statements are evidence that physical assault had taken place.

In Bryce's case the prosecutor asserts that the decision about his concern about which way she went home was "fatherly" or otherwise and whether his story of uninvolvement with Mary Watkins is true or whether her version of his part is true.

The decision is whether to believe a young girl's story, substantiated by three other witnesses or the story of two young men who admitted to chasing, sexual intercourse on various occasions.

DEFENCE'S SUMMING UP

Only evidence presented in court should be considered and the defendants should be acquitted if guilt is not certain. He says that the substantiated evidence about the girl's condition and behaviour is consistent with regret after the event rather than unwillingness for the act. He points out that Mary Watkins agreed to walk with Harrison and that three or four minutes of kissing took place. Harrison says this was followed by putting his penis between her legs and that such willingly indulged in behaviour might well have been tried immediately after. J.W.H.'s morals may well be reprehensible but that doesn't constitute rape or attempted rape which are serious crimes.

As for Bryce the decision is a straight forward one of whose story, the girl's or his, is to be believed according to Defence. He points

APPENDIX 8B (cont.)

out inconsistencies between her evidence here and in the Magistrate's court - who was with her when the bolt rattled, where, how and by whom she was hit. He suggested that if the girl's evidence seems fabricated or embellished the jury should not convict the men.

JUDGE'S SUMMING-UP

Rape means sexual intercourse with a girl or woman without her consent and by the use of force and fear. It is for you to decide whether Harrison and/or Bryce raped or attempted to rape Mary Watkins. Sexual intercourse means that there was at least partial penetration, full and complete intercourse is not necessary for this purpose. An attempt is something which falls short of the full offence. It is steps on the way to the commission of a crime which would be a crime if the act were completed. You have to be satisfied that there was the intention to commit rape before you can convict of an attempt.

The burden of proof lies on the crown from beginning to end, you must be satisfied that either of the accused is guilty before you can convict. The crown have to prove, beyond reasonable doubt, that the girl did not consent.

When the woman makes a complaint of a sexual attack it is dangerous to convict on the word of the woman, you must look for corroboration of her evidence from a source other than herself. Corroboration means some evidence independent of the woman's evidence. Nevertheless if you feel justified you can convict on her evidence even though it is not corroborated. You can take evidence of the accused as corroboration if it agrees with the story of the complainer. The vital matter on which you should seek corroboration is the question of consent.

I would now like you to reach a verdict on each of the four charges.

APPENDIX 8B (cont.)

out inconsistencies between her evidence here and in the Magistrate's court - who was with her when the bolt rattled, where, how and by whom she was hit. He suggested that if the girl's evidence seems fabricated or embellished the jury should not convict the men.

JUDGE'S SUMMING-UP

Rape means sexual intercourse with a girl or woman without her consent and by the use of force and fear. It is for you to decide whether Harrison and/or Bryce raped or attempted to rape Mary Watkins. Sexual intercourse means that there was at least partial penetration, full and complete intercourse is not necessary for this purpose. An attempt is something which falls short of the full offence. It is steps on the way to the commission of a crime which would be a crime if the act were completed. You have to be satisfied that there was the intention to commit rape before you can convict of an attempt.

The burden of proof lies on the crown from beginning to end, you must be satisfied that either of the accused is guilty before you can convict. The crown have to prove, beyond reasonable doubt, that the girl did not consent.

When the woman makes a complaint of a sexual attack it is dangerous to convict on the word of the woman, you must look for corroboration of her evidence from a source other than herself. Corroboration means some evidence independent of the woman's evidence. Nevertheless if you feel justified you can convict on her evidence even though it is not corroborated. You can take evidence of the accused as corroboration if it agrees with the story of the complainer. The vital matter on which you should seek corroboration is the question of consent.

I would now like you to reach a verdict on each of the four charges.

APPENDIX 8C

HARRISON - SUMMARY OF SORTING OF STATEMENTS BY 20 PERSONS

STATE-MENT	CLASS	SCALE VALUE	Q	ACCUMULATIVE FREQUENCY										
				A	B	C	D	E	F	G	H	I	J	K
				0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	10-11
1	I	6.8	2.4	1.00	1.00	.95	.95	.95	.90	.40	.35	.15	0	0
2	G	5.35	1.2	0	0	.05	.05	.35	.80	1.00	1.00	1.00	1.00	1.00
3	I	8.0	2.0	1.00	1.00	1.00	1.00	1.00	.90	.75	.50	.25	.10	0
4	I	7.5	3.0	1.00	1.00	1.00	.95	.85	.70	.60	.35	.15	.05	0
5	I	8.2	1.7	1.00	1.00	1.00	1.00	1.00	.95	.85	.55	.30	.10	.05
6	I	8.7	1.4	1.00	1.00	1.00	1.00	1.00	.95	.80	.75	.35	.15	0
7	G	5.0	2.5	0	0	.10	.25	.50	.70	.90	1.00	1.00	1.00	1.00
8	G	4.2	2.0	0	.05	.15	.45	.70	.80	.95	1.00	1.00	1.00	1.00
9	G	3.3	2.0	.05	.15	.40	.70	.85	.95	1.00	1.00	1.00	1.00	1.00
10	G	4.9	1.1	0	0	0	.10	.55	.95	1.00	1.00	1.00	1.00	1.00
11	G	5.0	1.75	0	0	.10	.25	.50	.80	.90	.90	.95	1.00	1.00
12	G	5.4	0.75	0	0	.10	.15	.25	.90	.95	.95	1.00	1.00	1.00
13	G	4.6	1.45	.05	.05	.10	.30	.65	.90	.95	.95	1.00	1.00	1.00
14	G	5.3	0.9	0	.05	.05	.15	.30	.95	1.00	1.00	1.00	1.00	1.00
15	G	4.1	2.1	0	0	.25	.45	.75	1.00	1.00	1.00	1.00	1.00	1.00
16	G	4.5	2.4	0	0	.25	.40	.65	.85	.95	.95	1.00	1.00	1.00
17	I	6.5	1.4	1.00	1.00	1.00	1.00	1.00	.85	.35	.15	.05	0	0
18	G	4.25	3.0	0	.10	.20	.45	.65	.70	.85	.95	1.00	1.00	1.00
19	I	6.5	0.6	1.00	1.00	1.00	1.00	1.00	.95	.15	.10	0	0	0
20	G	4.7	2.5	0	.10	.25	.40	.55	.90	.90	1.00	1.00	1.00	1.00
21	G	5.5	0.7	0	0	.05	.10	.20	.90	.95	1.00	1.00	1.00	1.00
22	G	5.0	2.3	0	.05	.20	.35	.50	.80	.90	.90	.95	.95	1.00
23	I	6.5	0.6	1.00	1.00	1.00	1.00	1.00	.95	.15	.15	.10	.05	0
24	I	8.3	1.5	1.00	1.00	1.00	1.00	1.00	1.00	.90	.60	.25	.15	0
25	G	4.2	1.6	0	.05	.20	.45	.85	.95	1.00	1.00	1.00	1.00	1.00

I = INNOCENT G = GUILTY

APPENDIX 8C (cont.)

HARRISON - SUMMARY OF SORTING OF STATEMENTS BY 20 PERSONS

				ACCUMULATIVE FREQUENCY										
STATE-MENT	CLASS	SCALE VALUE	Q	A 0-1	B 1-2	C 2-3	D 3-4	E 4-5	F 5-6	G 6-7	H 7-8	I 8-9	J 9-10	K 10-11
26	G	4.0	1.8	.05	.10	.25	.50	.80	1.00	1.00	1.00	1.00	1.00	1.00
27	G	5.4	1.1	0	0	0	.05	.35	.80	.90	1.00	1.00	1.00	1.00
28	G	5.4	0.7	0	0	0	.05	.20	.25	.90	.90	.90	1.00	1.00
29	I	6.6	0.6	1.00	1.00	1.00	1.00	.90	.85	.15	.10	.05	0	0
30	G	4.0	1.1	.05	.10	.10	.25	.50	.90	1.00	1.00	1.00	1.00	1.00
31	I	6.5	1.3	1.00	1.00	1.00	1.00	.95	.70	.30	.10	.05	0	0
32	I	8.0	2.9	1.00	1.00	1.00	1.00	1.00	1.00	.75	.50	.35	.15	0
33	G	2.8	2.8	.02	.03	.55	.70	.90	1.00	1.00	1.00	1.00	1.00	1.00
34(i)	G	4.0	2.7	.10	.20	.40	.50	.75	.95	1.00	1.00	1.00	1.00	1.00
(ii)	G	2.3	2.9	.10	.40	.65	.70	.80	.90	.95	1.00	1.00	1.00	1.00
(iii)	G	5.3	0.8	0	.05	.15	.20	.25	.90	1.00	1.00	1.00	1.00	1.00
35(i)	G	5.3	1.2	0	.10	.15	.20	.40	.90	1.00	1.00	1.00	1.00	1.00
(ii)	G	5.1	1.3	0	.10	.10	.20	.45	1.00	1.00	1.00	1.00	1.00	1.00
(iii)	G	5.8	1.5	0	.05	.10	.25	.55	.95	1.00	1.00	1.00	1.00	1.00
36	I	7.2	1.3	1.00	1.00	1.00	1.00	1.00	1.00	.55	.25	.10	.05	.05
37	I	6.5	0.6	1.00	1.00	1.00	1.00	.95	.90	.10	.10	.05	0	0
38	I	6.6	0.8	1.00	1.00	1.00	.95	.95	.85	.25	.10	0	0	0
39	I	7.7	2.1	1.00	1.00	1.00	1.00	.90	.80	.65	.40	.20	.05	0
40	I	6.9	2.2	1.00	1.00	1.00	1.00	.90	.85	.45	.35	.10	.85	0
41	I	6.5	0.6	1.00	1.00	1.00	.95	.95	.95	.15	.05	.05	0	0
42	I	6.9	1.8	1.00	1.00	1.00	.95	.85	.85	.45	.25	.15	.05	.05
43	G	5.4	1.3	0	0	0	.10	.35	.75	.95	1.00	1.00	1.00	1.00
44	G	5.5	0.8	0	0	0	.10	.20	.85	.95	.95	1.00	1.00	1.00
45	I	7.3	1.7	0	0	1.00	0	.95	.85	.55	.35	.05	.05	0
46	G	3.1	1.1	.15	.20	.45	.90	1.00	1.00	1.00	1.00	1.00	1.00	1.00

APPENDIX 8C (cont.)

HARRISON - SUMMARY OF SORTING OF STATEMENTS BY 20 PERSONS

STATE-MENT	CLASS	SCALE VALUE	Q	ACCUMULATIVE FREQUENCY										
				A 0-1	B 1-2	C 2-3	D 3-4	E 4-5	F 5-6	G 6-7	H 7-8	I 8-9	J 9-10	K 10-11
47	I	6.5	2.5	1.00	1.00	1.00	.90	.90	.50	.40	.25	.15	.10	0
48	G	4.4	3.3	0	.05	.20	.45	.55	.70	.85	.90	.95	1.00	1.00
49	G	4.6	3.0	0	.15	.25	.30	.60	.70	.90	.95	1.00	1.00	1.00
50	I	7.0	3.1	1.00	1.00	1.00	1.00	.95	.60	.50	.30	.15	0	0
51	G	3.0	3.5	0.10	.30	.50	.65	.70	.85	.95	1.00	1.00	1.00	1.00
52	G	4.25	4.2	0	.20	.40	.45	.60	.65	.80	.90	1.00	1.00	1.00
53	G	0.8	1.5	.55	.80	.85	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
54	G	1.7	2.9	.20	.60	.70	.75	.90	.95	1.00	1.00	1.00	1.00	1.00
55	G	5.5	0.5	0	0	0	0	.15	.95	.95	.95	.95	1.00	1.00
56	I	6.6	0.5	1.00	1.00	1.00	1.00	1.00	1.00	.10	.10	.10	.05	0
57	I	6.9	1.3	1.00	1.00	1.00	1.00	1.00	.95	.40	.20	.10	.05	0
58	I	6.6	0.7	1.00	1.00	1.00	1.00	1.00	1.00	.20	.15	.05	.05	0
59	I	6.5	0.6	1.00	1.00	1.00	1.00	1.00	.95	.10	.10	0.5	.05	0
60	G	5.4	0.8	0	0	.05	.20	.25	.85	.85	.95	1.00	1.00	1.00
61	G	4.8	1.3	0	0	.15	.20	.55	.85	.90	1.00	1.00	1.00	1.00
62	I	6.7	0.7	1.00	1.00	1.00	1.00	1.00	1.00	.25	.15	.10	0	0
63	I	6.6	0.6	1.00	1.00	1.00	1.00	1.00	.95	.15	.10	.05	0	0

APPENDIX 9A - QUESTIONS AND OUTCOMES GIVEN TO SUBJECTS IN THE "SPECIFIED" TREATMENT

A. What decision do/did you think the firemen will/would reach?

	OUTCOME	PROBABILITY
(1)	Firemen return to work & settle for 10% + reduced hours + productivity deal	
(2)	Firemen get 30% wage rise	
(3)	Firemen return to work & settle for 10% & reduced hours	
(4)	Firemen get 10% wage rise now & guarantee of substantial increases in Nov '78 & '79	
(5)	Compromise somewhere between 10 & 20% + reduced hours	

B. What do/did you think the part-time firefighters will/would do?

(1)	Carry on as they had been doing	
(2)	Come out on strike in support of the Firemen's claim	
(3)	Stop helping the troops because of F.B.U. pressure	

C. How long from Week 4 do/did you think the strike will/would go on?

(1)	More than 4 weeks (from week 4)	
(2)	More than 2 weeks (but less than 4)	
(3)	From 1-2 weeks longer	
(4)	Less than a week	

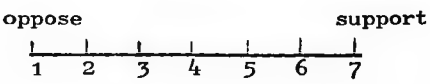
APPENDIX 9A (cont.)

D. What do/did you think the fire officers will/would do?

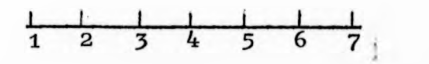
(1) Come out on strike for their own 40% wage claim	
(2) Carry on working and assisting troops & pursue their own claim by negotiation	
(3) Come out on strike in support of firemen's (F.B.U.) pay claim	
(4) Carry on as they are and wait and see what firemen get before pursuing their own pay claim	

E. How do/did you feel about :-

(1) The government pay policy



(2) The firemen's 30% claim



APPENDIX 9B - INSTRUCTIONS GIVEN TO SUBJECTS IN THE "SPECIFIED" AND "GENERATE" TREATMENTS FOR EACH OF THE THREE CONDITIONS

(a) "SPECIFIED" - Condition I

The instructions given to subjects in foresight (Condition I) in this treatment ran as follows:-

> This questionnaire is concerned with the current Firemen's dispute. You will find a brief summary of the events and related issues that have occurred over the past month.
>
> Once you have read this would you please follow the instructions outlined in the questionnaire sheet.
>
> Having read the description and drawing on your own knowledge of the situation would you please estimate the probabilities of the possible outcomes of this dispute. I would like you to give each outcome a probability value from 0 - 100%
>
> 0% - there is no chance of the outcome happening
> 100% - the outcome is certain

In the two hindsight conditions (6 weeks and 3 months after the strike had ended) the same subjects were given the following instructions:-

> As you remember, some time ago, while the firemen's strike was still on, you completed a questionnaire by providing probabilities for the occurrence of a number of possible outcomes to the dispute. The present interest is in the relation between the quality of your predictions and your ability to remember these predictions. For this reason, I would like you to fill out once again the same questionnaire, which you completed then, giving the same probabilities which you gave then. If you cannot remember the probability which you then assigned, give the probability which you would have given to each of the outcomes to the dispute.

Condition II. Subjects in this condition made only hindsight assessments to specified outcomes, their instructions ran as follows:-

> This questionnaire is concerned with the firemen's strike which has just ended. You will find a summary of the first four weeks of the strike and a summary of how it ended. Having read these would you please turn to the questionnaire sheet. On this I would like you to produce reconstructed probabilities - giving probabilities to the outcomes that you would have given during the strike. That is, try to think back to the middle of the strike, just before Christmas, and assign probabilities to the outcomes that you would have given then.

APPENDIX 9B (cont.)

In Condition III the above instructions were modified to exclude the reference to the "details" (i.e. summary of the first four weeks of the strike. After subjects had assigned likelihood assessments to the "specified" outcomes they were given the following instructions:-

> Having completed the first part of the questionnaire I would now like you to read a brief summary of the events of the first four weeks of the strike (7th Nov. to 5th Dec.). Having done this I would like you again to produce reconstructed probabilities for the various outcomes but in the light of the information given.

(b) "GENERATE" - Condition I

The instructions given to subjects in the foresight condition in this treatment were the same for the first paragraph in the "specified" treatment. After subjects had read the "details" they were given the following instructions:-

> Having read the description and drawing on your own knowledge of the situation would you please list a series of outcomes (either likely, unlikely or in-between). Against the outcome I would like you to estimate the likelihood of its occurrence. I would like you to give each outcome a probability value from 0 - 100%
>
> 0% - there is no chance of it happening
> 100% - it is certain to happen

In the two hindsight conditions (6 weeks and 3 months) subjects received similar instructions as those given to hindsight subjects in the "specified" treatment of Condition I. The instructions were modified to indicate that subjects had themselves generated the outcomes.

In Condition II subjects received the following instructions on the cover sheet of the questionnaire:-

> This questionnaire is concerned with the firemen's strike which has just ended. You will find a summary of the first four weeks and a summary of how it ended. What I would like you to do is to read this, then on the questionnaire sheet reconstruct a series of outcomes (either

APPENDIX 9B (cont.)

> likely, unlikely or in between) that you think may have occurred while the strike was still on. I would then like you to produce reconstructed probabilities - giving the probabilities that you would have given during the middle of the strike, that is, try to think back to the middle of the strike, just before Christmas, and assign probabilities you would have given then to the various outcomes.

Instructions given to subjects in Condition III on the cover sheet of the questionnaire were similar to those given immediately above. The instructions were modified to exclude reference to the "details". After generating outcomes and assigning likelihood assessments subjects were then instructed to read the "details" and, in the light of the new information reconstruct foresightful assessments to the outcomes generated in the first part of the questionnaire.

APPENDIX 9C - OUTCOMES PRODUCED BY SUBJECTS IN THE "GENERATE" TREATMENT AND THE CLASSIFICATION GIVEN TO THEM BY THE AUTHOR AND TWO JUDGES

Key: Numbers in last three right hand columns refer to the following classifications of outcomes:-

1. Settle for 10% + guarantee of rises in November 1978 and 1979.
2. Firemen get 30% wage rise.
3. Settle for 10% + reduced hours.
4. Compromise somewhere between 10% - 20%.
5. Not applicable.

SUB-JECT	OUTCOMES GENERATED	JUDGE 1	JUDGE 2	JUDGE 3 *
CONDITION I - FORESIGHT				
1	1. Victory to the firemen	2	2	2
	2. Victory to the government	3	3	3
	3. Victory to the government with a face saving formula for the firemen	1	4	4
	4. Strike peters out as individual firemen return to work	5	5	5
	5. Strike continues indefinitely	5	5	5
2	1. Firemen settle for 10% only	3	3	3
	2. Firemen settle for 10% + reduced hours + parity	1	1	1
	3. Firemen are offered more than 10% but less than maximum demand	4	4	4
	4. Firemen are offered and accept their maximum demands	2	2	2
3	1. Strike continues indefinitely	5	5	5
	2. Firemen accept new proposal	1	1	1
	3. Pay offer greater than 10%	4	4	4
	4. Further concessions leading to acceptance	4	4	5
4	1. Firemen accept less than 30%	4	4	4
	2. Government offer more than 10% + guarantee of parity	4	4	4
	3. Both parties will settle at a point midway between 10% and 20%	4	4	4

* Judge 3 was the author

APPENDIX 9C (cont.)

SUB-JECT	OUTCOMES GENERATED	JUDGE 1	JUDGE 2	JUDGE 3
5	1. Settle on terms of government offer	3	1	3
	2. Settlement on terms of F.B.U. i.e. full claim	2	2	2
	3. Compromise settlement of terms negotiated through C.A.B.	4	4	4
6	1. Government meets firemen's demands in full	2	2	2
	2. Collapse of strike in next few days	3	1	3
	3. Face-saving formula involving small over-run of 10% limit	4	4	4
7	1. Firemen refuse latest offer and hold out for face-saving formula	4	4	4
	2. Firemen accept latest proposals and strike ends	1	1	1
	3. Firemen refuse latest government offer and government capitulates	2	2	2
	4. Government meets all firemen's demands and strike ends within week	2	2	2
8	1. Accept 10% review	1	1	1
9	1. Firemen will agree to latest government offer	1	1	1
	2. Strike continues for few weeks more but eventually firemen give in	1	1	3
	3. Government will give in	2	2	2
10	1. Firemen's claim met in full	2	2	2
	2. Firemen awarded 10-15% + guaranteed pay increments	4	4	4
	3. Firemen awarded 10% + review + reduction in hours	4	1	1
	4. Strike fails, firemen receive only 10%	3	3	3
11	1. Firemen accept present offer i.e. 10% + review	1	1	1
	2. Firemen accept present offer + reduction in hours	4	3	3
	3. Offered 10% rise and accepted	4	4	4
12	1. Firemen get more than 10%	4	4	4
	2. Firemen settling for only 10%	3	3	3
	3. 10% now + review later	1	1	1

APPENDIX 9C (cont.)

SUB-JECT	OUTCOMES GENERATED	JUDGE 1	JUDGE 2	JUDGE 3
13	1. F.B.U. demand met in full without further developments	2	2	2
	2. Firemen settle on government offer of only 10%	3	3	3
	3. Government's offer of productivity deal will be accepted without further developments	1	1	1
	4. Firemen get increase offer beyond 10%	4	4	4
14	1. Firemen get what originally went on strike for	2	2	2
	2. Government make firemen special case	2	4	4
	3. Final wage settlement will not be what it should be i.e. average industrial wage + 10%	2	3	3
15	1. Firemen will settle for 10%	3	3	3
	2. Firemen will settle for 10% now + a later award	1	1	1
	3. Firemen will accept a national average linked wage in exchange for giving up right to strike	4	1	4
CONDITION II				
1	1. Settle in government guide lines	3	3	3
	2. Settle within government guide lines + further guaranteed rises	1	1	1
	3. Settle at a higher rate than 10%	4	4	4
2	1. Accept 10% now + parity later	1	1	1
	2. Accept just 10%	3	3	3
	3. Reject offer continue to press for original claim	2	5	5
3	1. Some men left their jobs as receiving no money while strike was on	5	5	5
	2. Fire Officers alienated by firemen	5	5	5
4	1. F.B.U. accept government 10% offer +	1	1	1
	2. Government make firemen special case and give them rise they demand	2	2	2

APPENDIX 9C (cont.)

SUB-JECT	OUTCOMES GENERATED	JUDGE 1	JUDGE 2	JUDGE 3
13	1. F.B.U. demand met in full without further developments	2	2	2
	2. Firemen settle on government offer of only 10%	3	3	3
	3. Government's offer of productivity deal will be accepted without further developments	1	1	1
	4. Firemen get increase offer beyond 10%	4	4	4
14	1. Firemen get what originally went on strike for	2	2	2
	2. Government make firemen special case	2	4	4
	3. Final wage settlement will not be what it should be i.e. average industrial wage + 10%	2	3	3
15	1. Firemen will settle for 10%	3	3	3
	2. Firemen will settle for 10% now + a later award	1	1	1
	3. Firemen will accept a national average linked wage in exchange for giving up right to strike	4	1	4
CONDITION II				
1	1. Settle in government guide lines	3	3	3
	2. Settle within government guide lines + further guaranteed rises	1	1	1
	3. Settle at a higher rate than 10%	4	4	4
2	1. Accept 10% now + parity later	1	1	1
	2. Accept just 10%	3	3	3
	3. Reject offer continue to press for original claim	2	5	5
3	1. Some men left their jobs as receiving no money while strike was on	5	5	5
	2. Fire Officers alienated by firemen	5	5	5
4	1. F.B.U. accept government 10% offer +	1	1	1
	2. Government make firemen special case and give them rise they demand	2	2	2

APPENDIX 9C (cont.)

SUB-JECT	OUTCOMES GENERATED	JUDGE 1	JUDGE 2	JUDGE 3
5	1. Accept 10% rise	3	3	3
	2. 10% + guarantee of parity	1	1	1
	3. Between 10 - 15%	4	4	4
	4. Over 15%	4	4	4
	5. 10% + promise that firemen not subject to future pay restraint	1	1	1
6	1. Accept improved offer	1	1	1
	2. Public opposition against firemen - accept only 10%	3	3	3
	3. T.U.C. offer support to firemen, government give in	2	2	2
7	1. F.B.U. accept offer of week 9	1	1	1
	2. F.B.U. backing down without immediate rise	3	3	5
8	1. 10% + reduction in hours + review	1	1	1
	2. Government give in and allow 30%	2	2	2
	3. Concession by government 20% wage increase + review	4	4	4
	4. T.U.C./government confrontation	5	5	5
9	1. Union go back without substantial increases	1	1	1
	2. Firemen would gain more	2	5	4
	3. Strike go on for longer than 9 weeks	5	5	5
10	1. Take first offer	3	3	3
	2. Refuse all offers except one asking for	2	5	2
	3. Take 10% + review '78 and '79	1	1	1
	4. Take offer of January	1	1	1
11	1. Total defeat of firemen, i.e. only 10%	3	3	3
	2. 12½% + reduced hours	4	4	4
	3. Parity with skilled workers by 1979 + 10% now	1	1	1
	4. 30% rise	2	2	2
	5. 10% + reduction in hours	3	1	3
	6. Award in excess of 20%	4	2	4

APPENDIX 9C (cont.)

SUB-JECT	OUTCOMES GENERATED	JUDGE 1	JUDGE 2	JUDGE 3
12	1. 10% rise in line with government guide-lines	3	3	3
	2. 10% + guarantee of review in '78 & '79	1	1	1
	3. 30% to F.B.U.	2	2	2
	4. 10% + free accommodation	3	5	5
13	1. Settle for 10% only	3	3	3
	2. Intermediate settlement	4	4	4
	3. Shorter hours + 10%	4	3	3
	4. 10% now + promise of more later	1	1	1
	5. Push for higher claim	5	5	2
14	1. Government offer more 10%	4	4	4
	2. 10% + reduced hours	4	1	3
	3. 10% only + review	1	1	1
	4. Union holds out for more 10%	4	5	4
15	1. Firemen accept offer by week 9	1	1	1
CONDITION III				
1	1. Strike would be successful and government would give in	2	2	2
	2. Compromise - 10% + reduced hours	4	3	1
	3. Strike unsuccessful and firemen go back to work	1	3	3
2	1. F.B.U. get 15-20% wage rise	4	4	4
	2. Firemen help in emergencies	5	5	5
3	1. Offer 15% accepted	4	4	4
	2. Accept original government offer	3	3	3
	3. Accept original government offer + parity	1	1	1
	4. Offer 25%	4	2	2
4	1. Rejection of proposal	5	5	5
	2. Accept proposal	1	1	1

APPENDIX 9C (cont.) .

SUB-JECT	OUTCOMES GENERATED	JUDGE 1	JUDGE 2	JUDGE 3
5	1. Get what asking for	2	2	2
	2. Reduction in hours + 10%	1	1	1
	3. Strike lasting 2-3 weeks longer than it did	5	5	5
6	1. End of strike on firemen's terms	2	2	2
	2. End of strike on employers' terms	1	3	1
7	1. Settle in 10% guideline but no guarantee future wage rises	3	3	3
	2. Settle for 10% + guarantee better pay prospects	1	1	1
	3. 10% + fiddles, e.g. different grading structure, increased responsibility allowance	4	3	5
	4. Above 10%	4	4	4
	5. Settle at firemen's original demand	2	2	2
8	1. Accept 10% instead of 30%	1	1	1
	2. Government gives 30%	2	2	2
	3. 10% settlement accepted by firemen before week 5	5	3	5
9	1. Offer over 10%	4	4	4
	2. Accept 10% offer	1	1	1
	3. T.U.C. would call general strike causing government to call general election	5	5	5
10	1. Claim would be met in full	2	2	2
11	1. Strike continue until original demands met	2	2	2
	2. Accept 10% and go back to work	1	1	1

APPENDIX 11A - INFORMATION SHEET GIVEN TO SUBJECTS IN FORESIGHT CONDITION

The General Election of October 1974

This was the last General Election to be held. It was predicted that Labour would win with a fairly massive majority, but in the end the result proved to be close.

The outcome was as follows:-

Party	Number of seats	Percent share of total votes cast
Labour	319	39.2%
Conservative	277	35.8%
Liberal	13	18.3%
Scottish Nationalist Party	11	2.9%
National Front	nil	0.4%
Others	15	3.4%
Total	365	Total 100%

This result gave Labour an overall majority of 3 seats.

Over the whole country Labour achieved a swing of 2.2% from Conservative.

The overall turnout was 72.8% (this is the percentage of voters who actually voted).

The forthcoming General Election

The state of the parties at going to the polls is as follows:-

Party	Number of seats
Labour	306
Conservative	281
Liberal	13
Scottish Nationalist Party	11
National Front	0
Others	20
Speaker and Deputies	4
Total	365

APPENDIX 11A (cont.)

Bye-elections held since the last General Election have resulted in swings to Conservative. Those held in 1975, '76 and '77 resulted in an average swing to the Conservatives of 11.8%, those held in 1978 have resulted in an average swing of 7.1% to Conservative from Labour.

The Result of the General Election of 3rd May 1979

In the run up to the General Election the Opinion Polls consistently predicted a Conservative majority. This proved to be the case. The final outcome was as follows:-

Party	Number of seats	Percent share of total votes cast
Labour	268	36.9%
Conservative	339	43.9%
Liberal	11	13.8%
Scottish Nationalist Party	2	1.6%
National Front	nil	0.6%
Others	15	3.2%
Total	635	Total 100%

This result gave the Conservatives an overall majority of 43 seats.

Over the whole country Conservative achieved a swing of 5.2% away from Labour.

The Liberals lost three seats, there being a 2.5% swing against them.

The National Front slightly increased their share of the vote; in 1974 they polled 0.4% of the votes cast, they polled 0.6% of the votes cast in May 1979.

The Scottish Nationalist Party lost 9 seats, holding on to only 2.

The overall turnout at the General Election was 76% (this is the percentage of voters who actually voted).

APPENDIX 11C - QUESTIONNAIRE GIVEN TO LABOUR & CONSERVATIVE SUBJECTS IN THE HINDSIGHT CONDITION

1(a) What did you think the result of the May 1979 General Election would be?

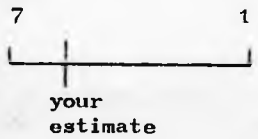
Would you please indicate this by estimating how likely you thought each of the following outcomes were, i.e. for each of the five possible outcomes would you indicate how likely you thought each was. Make your likelihood estimates in terms of percentages and ensure that the sum of your estimates total 100%.

OUTCOME	LIKELIHOOD
1. Overall majority[+] to Labour - i.e. a Labour Government.	
* 2. Overall majority to Conservative - i.e. a Conservative Government.	
3. Majority[++] to Labour (but not an overall one) - leading to Labour Government by forming some kind of pact with a minor party.	
4. Majority to Conservative (but not an overall one) - leading to a Conservative Government by forming some kind of pact with a minor party.	
5. Other outcome, e.g. Liberal government, National Front Government.	

[+] An overall majority is where one party has more seats than all the other parties put together.

[++] A majority is when one party has more seats than any other party but does not have an overall one.

In the rest of this questionnaire you will find that you are asked to estimate the likelihood of possible outcomes on a 7-point scale. To indicate what you think put a vertical line along the scale such that it corresponds to your estimate. Please do this for each scale provided. For example:-



APPENDIX 11C (cont.)

1(b) Would you indicate below for the same outcomes how surprised you would have been if it had occurred. Do this on the 7-point scales associated with each outcome.

1. Overall majority to Labour i.e. a Labour Government.

7 — 1
SURPRISED — NOT SURPRISED

* 2. Overall majority to Conservative - i.e. a Conservative Government.

7 — i
SURPRISED — NOT SURPRISED

3. Majority to Labour (but not an overall one) - Leading to Labour Government by forming some kind of pact with a minor party.

7 — 1
SURPRISED — NOT SURPRISED

4. Majority to Conservative (but not an overall one) - Leading to a Conservative Government by forming some kind of pact with a minor party.

7 — 1
SURPRISED — NOT SURPRISED

5. Other outcome e.g. Liberal Government, National Front Government

7 — 1
SURPRISED — NOT SURPRISED

2 At the October 1974 General Election the turnout was 72.8%. How large did you think the turnout would be at the May 1979 General Election? Would you indicate this by assessing the likelihood of the following possibilities:

1. Greater than 78%

7 — 1
VERY LIKELY — NOT VERY LIKELY

* 2. Between 73 - 78%

7 — 1
VERY LIKELY — NOT VERY LIKELY

3. About the same as before, i.e. between 71 - 73%

7 — 1
VERY LIKELY — NOT VERY LIKELY

APPENDIX 11C (cont.)

4. Between 68 - 71%

5. Less than 68%

3 What did you think the swing at the May 1979 General Election would be? For either one of the major parties to have won the General Election there would have had to have been a swing in votes to that party. Would you please indicate on the 7-point scales how likely you thought each of the swings listed below were:

1. Swing to Labour of more than 10%

2. Swing to Labour of between 5 - 10%

3. Swing to Labour of between 2 - 5%

4. Swing to Labour of between 0.1-2%

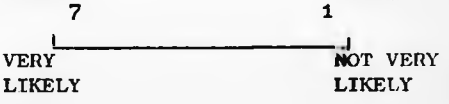
5. No swing to either Labour or Conservative

6. Swing to Conservative of between 0.1 - 2%

7. Swing to Conservative of between 2 - 5%

* 8. Swing to Conservative of between 5 - 10%

9. Swing to Conservative of more than 10%



APPENDIX 11C (cont.)

4 Swings to Labour or Conservative have, of course, consequences for the number of seats each party ends up with. However, there can be a disparity between the size of the national swing and what happens locally. Thus the number of seats predicted from a certain swing may not be the same as that actually achieved. For this question would you estimate how likely you thought the lead in seats (if any) was likely to be. As a rough guide the swing needed to achieve this is given in brackets.

Estimate how likely you thought the following majorities (not overall majorities) were on the 7-point scales provided for the possible outcomes listed below.

1. Labour majority of greater than 150 seats (about 3% swing to Lab)

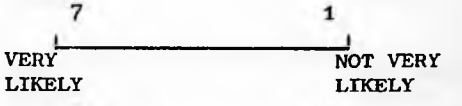
2. Labour majority of between 100 - 150 seats (about 2-3% swing to Labour)

3. Labour majority of between 40 - 100 seats (about 0 - 1% swing to Labour)

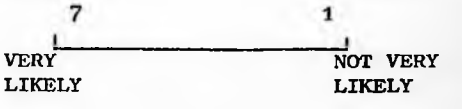
4. Labour majority of between 0 - 40 seats (about 0 - 1% swing to Conservative)

5. Conservative majority of between 0 - 40 seats (about 1 - 3% swing to Conservative)

* 6. Conservative majority of between 40 - 90 seats (about 3 - 5% swing to Conservative)



7. Conservative majority of greater than 90 seats (about 5% swing or more to Conservative)



APPENDIX 11C (cont.)

5 Three questions concerned with how you thought the LIBERALS would fare at the May 1979 General Election.

(a) At the General Election in October 1974 the Liberals polled 18.3% of the votes (about 5.5 million votes). What percentage did you think they would poll in the May 1979 General Election? Indicate this by estimating how likely you thought each of the following outcomes were:

1. Poll more than 25% of the votes

7 — VERY LIKELY — 1 — NOT VERY LIKELY

2. Poll between 19 - 25% of the votes

7 — VERY LIKELY — 1 — NOT VERY LIKELY

3. Poll about the same (17 - 19%)

7 — VERY LIKELY — 1 — NOT VERY LIKELY

* 4. Poll between 12 - 17% of the votes

7 — VERY LIKELY — 1 — NOT VERY LIKELY

5. Poll less than 12% of the votes

7 — VERY LIKELY — 1 — NOT VERY LIKELY

(b) How likely did you think the following swings for Liberal were?

1. More than 5% to Liberal

7 — VERY LIKELY — 1 — NOT VERY LIKELY

2. Between 3 - 5% to Liberal

7 — VERY LIKELY — 1 — NOT VERY LIKELY

3. Between 0.1 - 3% to Liberal

7 — VERY LIKELY — 1 — NOT VERY LIKELY

4. No swing

7 — VERY LIKELY — 1 — NOT VERY LIKELY

* 5. Between 0.1 - 3% away from Liberal

7 — VERY LIKELY — 1 — NOT VERY LIKELY

6. Between 3 - 5% away from Liberal

7 — VERY LIKELY — 1 — NOT VERY LIKELY

7. More than 5% away from Liberal

7 — VERY LIKELY — 1 — NOT VERY LIKELY

APPENDIX 11C (cont.)

(c) National swings might not accurately reflect seats lost or gained. This being the case would you estimate how likely you thought the following number of seats for Liberal were. (At the time of going to the polls Liberal had 14 seats). As a rough guide the swing needed to achieve these seat numbers is given.

1. More than 30 seats (swing of more than 5% to Liberal)

2. Between 23 - 30 seats (swing of 3 - 5% to Liberal)

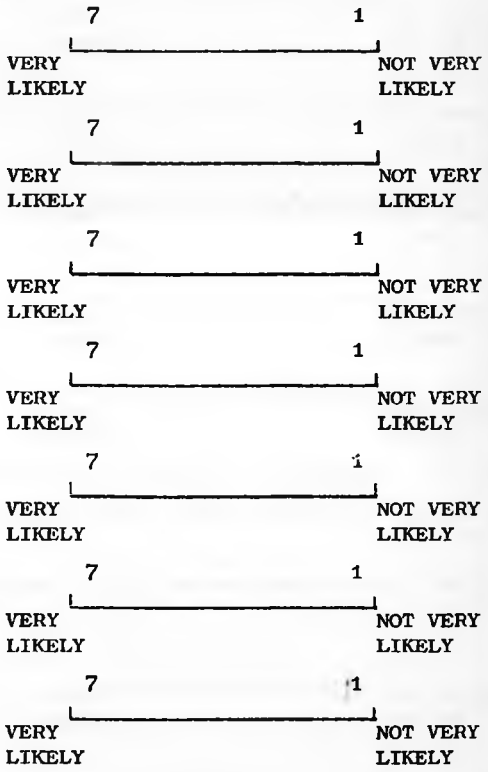
3. Between 14 - 23 seats (swing of 1 - 3% to Liberal)

4. The same number of seats (no swing either way)

* 5. Between 6 - 12 seats (swing of 1 - 3% away from Liberal)

6. Between 4 - 6 seats (swing of 3 - 5% away from Liberal)

7. Less than 4 seats (swing of more than 5% away from Liberal)



APPENDIX 11C (cont.)

6 THE NATIONAL FRONT At the General Election of October 1974 the National Front polled 0.4% of the votes cast (this represents about 140,000 votes). How well did you think they would fare in the May 1979 General Election? Would you indicate this by estimating how likely you thought each of the following outcomes were.

1. Poll more than 1% of the votes cast

* 2. Poll between 0.4 - 1% of votes cast

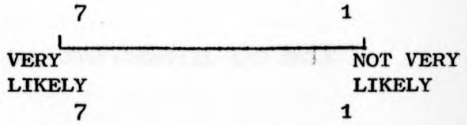
3. Poll about the same (0.4%)

4. Poll between 0.2 - 0.4% of votes cast

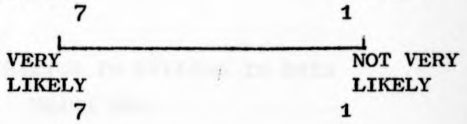
5. Poll less than 0.2% of votes cast

7 THE SCOTTISH NATIONALIST PARTY At the General Election of October 1974 the Scottish Nationalist Party had 11 seats. How many did you think they would get in May 1979?
Would you indicate this by estimating how likely you thought each of the following outcomes were.

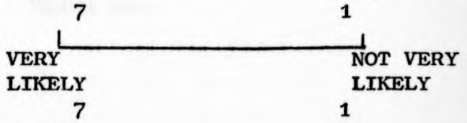
1. More than 20 seats (swing of more than 4% to the SNP)



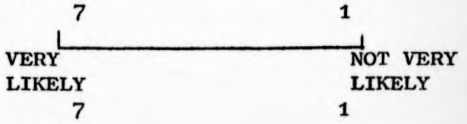
2. Between 13-20 seats (swing of about 3% to SNP)



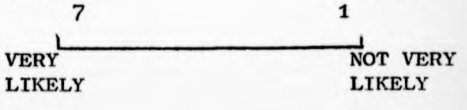
3. Between 10-12 seats, i.e. about the same



4. Between 6-10 seats (swing of about 2% away from SNP)



* 5. Less than 6 seats (swing of about 3% away from SNP)



APPENDIX 11C (cont.)

8 Which party did YOU want to see in office (i.e. form a Government) following the General Election? Please tick one of the following as appropriate:- ++

(a) Conservative
(b) Labour
(c) Liberal
(d) Other (please specify)

9 Who did you vote for in the General Election? Please tick one of the following as appropriate: ++

(a) Conservative
(b) Labour
(c) Liberal
(d) Other (please specify)

++ THIS INFORMATION WILL BE TREATED AS STRICTLY CONFIDENTIAL AND ONLY USED FOR RESEARCH PURPOSES.

THANK YOU VERY MUCH FOR YOUR TIME AND CO-OPERATION IN FILLING IN THIS QUESTIONNAIRE - IT IS VERY MUCH APPRECIATED. THANK YOU.

APPENDIX 12 - DESCRIPTIVE AND INFERENTIAL STATISTICS

Experiments investigating hindsight bias/creeping determinism reported in this thesis have all, with the exception of Experiment 8, used non-parametric statistics. There were two sets of reasons for this: those concerned with (1) assumptions underlying the use of parametric tests; (2) the distribution of responses subjects gave in the various experimental treatments.

Parametric statistics, such as the t test and analysis of variance, whilst estimating population parameters also make three major assumptions about the population from which the data was drawn, and the data itself. These assumptions are that (1) the samples were drawn randomly from the population; (2) the data should be normally distributed; and (3) the data should be at least on a true interval scale. Some comment on each will show why non-parametric statistics have been used in this thesis.

The first assumption could be met by the experiments in question. In all these experiments samples were randomly drawn, within particular specified constraints, and subjects were randomly allocated to experimental treatments.

The second assumption, that scores should be normally distributed could not, in general, be met. This was for two reasons, (1) the number of subjects in each experimental treatment was between 10 and 15. With such numbers one would hardly expect the responses to be normally distributed. On its own this problem would not be too severe but does become so in light of the second reason. (2) The subjective likelihood assessments given by different people to the same scenario and outcome in the same experimental treatment showed great variation. On some occasions the range of assessments given by subjects asked to express their likelihoods in percentages was 0 to 100. This merely demonstrates that opinions differed as to the likelihood of outcomes. Hence with such a range of responses from a small sample the data could not be considered as being normally distributed, but as skewed.

The third and final assumption, that the data should be on a true interval scale, could not be met either. In the experiments where

APPENDIX 12 (cont.)

subjects were asked to make assessments in terms of percentages (Experiments 1, 2, 3, 5 and part of 6) it might appear that the data was on an interval scale. However, it became apparent, from asking subjects and looking at patterns of responses, that this was not so.

Two reasons can be cited: (1) if one subject gave a response of 30% and another a response of 60%, it did not imply, as is required, that the latter saw the outcome as twice as likely as the former. We need to have some indication that such a relationship would hold for it to be classified as interval scale data. (2) In one experiment (Experiment 5) subjects were not instructed to ensure that the percentages for each of the outcomes to a question totalled 100%. Adding assessments to a number of outcomes to a question showed some to be in excess of this figure. Given this we can hardly assume that an assessment of 30% by a subject whose percentages totalled 100% meant that outcome was half as likely when a subject gave an assessment of 60% but the total was in excess of 100%. For these two reasons, then, the data cannot be considered as being on a true interval scale and must be regarded as ordinal.

The second set of reasons for using non-parametric statistics concerns, as indicated earlier, the distribution of likelihood assessments given by subjects. As indicated above there was often quite a range of responses. To compute means and use parametric statistics would have given undue weight to the extreme values. This is especially true if subjects gave responses within a relatively narrow range and one or two gave extreme responses. The mean would be unrepresentative in such cases, particularly where N is small. (I am grateful to McCauley & Jaques, 1979 for this argument).

The above two sets of reasons meant that medians and non-parametric statistics were appropriate in the experiments being discussed here.

To summarise, the assumptions of a normal distribution and the data being on a true interval scale could not be met. Furthermore, the distribution of responses indicated that parametric statistics were appropriate as they are less sensitive to the presence of a few extreme scores than parametric statistics.

APPENDIX 13 - SAMPLE SIZE

Two of the experiments (Experiment 5 and 6) reported in this thesis have been accepted for publication in the British Journal of Social and Clinical Psychology. In the course of the review process and from personal communications with Fischhoff one comment has appeared consistently. This has centred round the suggestion that the studies have low power due to relatively small sample size. Simply stated the problem is, according to Fischhoff (personal communication), "the chances of detecting significant hindsight effects diminishes with small sample sizes." Fischhoff's research, typically, used between 20 and 30 subjects per treatment, that reported in this thesis used between 10 and 15 per treatment.

The above problem raised by the reviewers and Fischhoff is not a simple one, however. We not only want to know how reliable an effect is (here hindsight bias/creeping determinism), but also, and just as importantly, how strong the effect is. The two issues are related because we are only interested in reliable effects if they are also reasonably strong ones.

To appreciate the issues more fully a general discussion of sample size, significant differences and the strength of an effect will first be given. This will be followed by a consideration of how these issues relate to the experiments reported in this thesis.

In estimating the number of subjects needed in an experiment two major factors need to be considered: (1) how big a difference between experimental treatments is considered important; (2) whether the research is concerned with replication or extension of previous findings.

The first factor refers to what is technically known as "strength of association" (Keppel, 1973; Linto & Gallo, 1975; Hays, 1973). The significance level obtained from the application of a statistical test tells us no more than a relationship between variables exists at a specified level of probability. It follows that although the relationship is a real one it might also be a weak one. The strength of association between the experimental treatments and resultant behaviour tells us how strong or weak this relationship is.

APPENDIX 13 - SAMPLE SIZE

Two of the experiments (Experiment 5 and 6) reported in this thesis have been accepted for publication in the British Journal of Social and Clinical Psychology. In the course of the review process and from personal communications with Fischhoff one comment has appeared consistently. This has centred round the suggestion that the studies have low power due to relatively small sample size. Simply stated the problem is, according to Fischhoff (personal communication), "the chances of detecting significant hindsight effects diminishes with small sample sizes." Fischhoff's research, typically, used between 20 and 30 subjects per treatment, that reported in this thesis used between 10 and 15 per treatment.

The above problem raised by the reviewers and Fischhoff is not a simple one, however. We not only want to know how reliable an effect is (here hindsight bias/creeping determinism), but also, and just as importantly, how strong the effect is. The two issues are related because we are only interested in reliable effects if they are also reasonably strong ones.

To appreciate the issues more fully a general discussion of sample size, significant differences and the strength of an effect will first be given. This will be followed by a consideration of how these issues relate to the experiments reported in this thesis.

In estimating the number of subjects needed in an experiment two major factors need to be considered: (1) how big a difference between experimental treatments is considered important; (2) whether the research is concerned with replication or extension of previous findings.

The first factor refers to what is technically known as "strength of association" (Keppel, 1973; Linto & Gallo, 1975; Hays, 1973). The significance level obtained from the application of a statistical test tells us no more than a relationship between variables exists at a specified level of probability. It follows that although the relationship is a real one it might also be a weak one. The strength of association between the experimental treatments and resultant behaviour tells us how strong or weak this relationship is.

APPENDIX 13 (cont.)

The strength of association measure is in terms of the amount of variance accounted for, and is sometimes estimated by the use of the following general formula (Linton & Gallo, 1975, p.332):

$$\hat{r}_m^2 = \frac{Q^2}{Q^2 + N}$$

where $\hat{r}_m^2$ = estimate of the strength of association
Q = numerical value of the inferential statistic
N = the sample size

Where analysis of variance is used strength of association is calculated using omega squared (w^2) (Hays, 1963).

The point about this index, whichever formula is used to calculate it, is that sample size exerts a great influence upon it. As Rosenthal & Gaito (1963) show, if two experiments are run, one with an N of 5 and one with an N of 20 and both yield a significant result at P = 0.05 it is the one with the smaller sample size that yields greater strength of association. With sample size constant the smaller the value of the greater the strength of association.

Ideally, psychological research should aim at obtaining reliable differences which also account for a reasonable proportion of the variance (10% or more is the figure usually accepted). However, under certain circumstances we may find these two criteria to be in conflict. This happens because there is a trade off between the power of a test (Cohen, 1962, 1969) and strength of association between variables.

The power of a statistical test refers to the probability of rejecting the null hypothesis when it is false. It is calculated using β , which gives the probability of committing a Type II error. Generally, the probability of committing a Type II error decreases as sample size increases, hence increase in sample size increases the power of a statistical test.

APPENDIX 13 (cont.)

We are now in a position to consider the above discussion in terms of hindsight bias/creeping determinism. As mentioned earlier, one potential problem with the studies reported in this thesis is their low power. By using small sample sizes relative to those in Fischhoff's original studies one may reduce the chances of detecting significant hindsight effects if they exist. Tversky & Kahneman's (1971) "Law of Small Numbers" deals in detail with this problem.

This is only a problem, however, if we are dealing with an effect which is relatively weak (i.e. accounts for only a small proportion of the variance) in the first place. It is not possible to determine the size of the hindsight effect from published research. It will be recalled from Chapter 3 that the work of Fischhoff and his colleagues was criticised on the grounds that, among other things, their presentation of data and statistical analysis provided little or no insight into how strong an effect hindsight bias/creeping determinism is. The main argument for this (see p.) came from the criticism that they failed, on the whole, to provide inferential statistical treatment to individual experimental conditions.

In designing and analysing the experiments reported in this thesis it was originally assumed that the hindsight effect was a reasonably strong one. Hence sample size, within certain limits, was not considered to be a methodological problem. It soon became apparent, as the results in the early experiments (which were actually designed and run before the ones appearing later in this thesis) indicated, was not as strong as researchers seem to have assumed.

The assumption was, then, that using 10 to 15 subjects in each experimental treatment, as most of the experiments reported here do, would provide adequate statistical power and, more importantly, give a better indication of the strength of the effect. Using such numbers it was further assumed that effects due to outcome knowledge, if found, would be interesting ones and ones which might apply outside of the laboratory.

APPENDIX 13 (cont.)

In summary, the issue of sample size is an important consideration in the context of the experiments reported in this thesis. If the concern was solely replication of Fischhoff's results then the number of subjects in each treatment would have had to have been considerably increased. However, as the interest, overall, was with the magnitude of the effect smaller sample sizes are justified.

The statistical tests used to investigate foresight/hindsight differences have, throughout this thesis, been non-parametric. Unfortunately, as Linton & Gallo (1975) point out: "little attention has been paid to developing strength of association measures for ordered data" (p.333). Such formulae exist for the Kruskal-Wallis test, Friedman test and the Rank-sums test, but not for the non-parametric tests used in this thesis.

564

END